



**HAL**  
open science

# Sparse Representation-Based Classification of Mysticete Calls

Thomas Guilment, François-Xavier Socheleau, Dominique Pastor, Simon Vallez

► **To cite this version:**

Thomas Guilment, François-Xavier Socheleau, Dominique Pastor, Simon Vallez. Sparse Representation-Based Classification of Mysticete Calls. *Journal of the Acoustical Society of America*, 2018, 144 (3), pp.1550. 10.1121/1.5055209 . hal-01891652

**HAL Id: hal-01891652**

<https://imt-atlantique.hal.science/hal-01891652v1>

Submitted on 9 Oct 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sparse Representation-Based Classification of Mysticete Calls

Thomas Guilment,<sup>\*</sup> Francois-Xavier Socheleau,<sup>†</sup> and Dominique Pastor<sup>‡</sup>

*IMT Atlantique, Lab-STICC, Bretagne Loire University,  
Technopole Brest-Iroise CS83818, Brest 29238, France*

Simon Vallez<sup>§</sup>

*Sercel, 12 Rue de la Villeneuve, 29200 Brest, France*

(Dated: October 9, 2018)

## Abstract

This paper presents an automatic classification method dedicated to mysticete calls. This method relies on sparse representations which assume that mysticete calls lie in a linear subspace described by a dictionary-based representation. The classifier accounts for noise by refusing to assign the observed signal to a given class if it is not included into the linear subspace spanned by the dictionaries of mysticete calls. Rejection of noise is achieved without feature learning. In addition, the proposed method is modular in that, call classes can be appended to or removed from the classifier without requiring retraining. The classifier is easy to design since it relies on a few parameters. Experiments on five types of mysticete calls are presented. It includes Antarctic blue whale Z-calls, two types of “Madagascar” pygmy blue whale calls, fin whale 20 Hz calls and North-Pacific blue whale D-calls. On this dataset, containing 2185 calls and 15000 noise samples, an average recall of 96.4% is obtained and 93.3% of the noise data (persistent and transient) are correctly rejected by the classifier.

PACS numbers: PACS: 30.Sf, 30.Wi

---

<sup>\*</sup> [thomas.guilment@imt-atlantique.fr](mailto:thomas.guilment@imt-atlantique.fr); Corresponding author.

<sup>†</sup> [fx.socheleau@imt-atlantique.fr](mailto:fx.socheleau@imt-atlantique.fr)

<sup>‡</sup> [dominique.pastor@imt-atlantique.fr](mailto:dominique.pastor@imt-atlantique.fr)

<sup>§</sup> [simon.vallez@sercel.com](mailto:simon.vallez@sercel.com)

## 1 I. INTRODUCTION

2 Passive acoustic monitoring (PAM) is very useful tool for helping scientists study marine mam-  
3 mals [1], detect their presence during seismic surveys and as a consequence, mitigate the impact  
4 of man-made acoustic activities [2, 3]. The success of PAM has led to an increasing deployment  
5 of underwater acoustic recorders across many oceans [4]. As a result, the development of efficient  
6 and robust automatic methods is needed to analyze the growing amount of acoustic data gener-  
7 ated by these recording systems. Such methods are helpful for human analysts to detect, classify,  
8 locate, track or count marine mammals.

9 PAM is particularly relevant for mysticetes or baleen whales which are known to produce a  
10 wide variety of underwater sounds [5–7]. Their repertoire is composed of tonal [8, 9], frequency-  
11 modulated (FM) [10], pulsive [11, 12] sounds and other calls with exotic names such as boings  
12 [13], moans and grunts [14], exhalation and gunshot [15], and “star-wars” vocalization [16]. Mys-  
13 ticete calls exhibit different levels of variability. Some calls, such as Antarctic blue whale Z-calls  
14 [17], only show slight inter-annual and seasonal variations [8], whereas other vocalizations, such  
15 as songs produced by bowhead whales [3, 18], fully change from one year to another [19]. In  
16 between, there are a variety of calls with the same signal structure but with parameters, such as  
17 duration and/or bandwidth and/or FM rate, whose values may change over time [7].

18 Automatic classifiers of mysticete calls face several challenges. As any pattern recognition  
19 algorithms, they have to identify the salient features of the calls of interest. However, this may  
20 be difficult because (i) signal-to-noise ratios can be low, (ii) propagation effects can distort the  
21 call features [20] and, (iii) the selected features must not only describe and discriminate the calls  
22 of interest, but also [21] “*provide contrast to any other type of signal that is likely to occur*”  
23 in the same acoustic context. Past experiments have shown that acoustic recordings can contain  
24 a wide variety of interfering transient sounds in the frequency range of mysticete calls [22–26].  
25 Therefore, providing classifiers with a rejection option that refuses to assign a signal of no interest  
26 to any class is of prime importance for PAM applications.

27 In the context of multiclass classification, most automated techniques for mysticete calls im-  
28 plement a two-step procedure. They usually operate in the frequency or cepstral domain and first  
29 extract sound attributes like start frequency, end frequency, frequency slope, duration etc. A super-  
30 vised learning algorithm then maps these attributes to a call class after learning training examples  
31 labeled by human analysts. Classifier of this kind include aural classification [27], neural networks

32 [3], hidden Markov models [28], quadratic discriminant function analysis [29], Gaussian mixture  
33 models [30] or classification trees [31]. More recently, Halkias *et al.* [25] proposed an alternative  
34 approach based on hybrid generative/discriminative models commonly used in machine learning.  
35 This method involves injecting a spectrogram image of the sound to process into a multiple-layer  
36 neural network. The main advantage of the used network is that it automatically learns the signal  
37 attributes from unlabeled data and does not rely on “hand-engineered” features.

38 Although applied with success in specific contexts, state-of-the-art methods may however show  
39 some limitations. For instance, some classifiers lack of general applicability because they are tuned  
40 for specific species. This is the case of spectrogram correlation [32], non-spectrogram correlation  
41 [13], vector quantization algorithm and dynamic time warping [33]. Others may require to tune  
42 many (hyper)-parameters [25, 29]. In case these parameters are not easy to physically interpret,  
43 their numerical values may be difficult to set, which can limit the robustness of the classifier or lead  
44 to under- or over-fitting. Moreover, some methods offer a rejection option that rely on parametric  
45 models of noise [24] or require the classifier to learn the features of the unwanted signals [25].  
46 Exhaustive noise learning or modeling is hardly feasible in practice since the underwater acoustic  
47 environment is very complex and contains many transient signals with very different features.  
48 In addition, these features may fluctuate in time and space so that they may greatly vary from  
49 one dataset to another. Finally, most existing classifiers lack of modularity/flexibility and are  
50 often designed for a specific set of calls, so that adding or removing a call class usually requires  
51 to “retrain” the entire classifier. In a PAM context, where the same classifier may be used on  
52 platforms operating at different geographic locations and at different time of the year, offering the  
53 capability of selecting online the class of calls taken into account by the classifier may have an  
54 operational interest. Classes corresponding to species whose habitats are known to be far away  
55 from the sensor may therefore be removed from the classifier, thus reducing the probability of  
56 miss-classification.

57 In this paper, a general method capable of classifying multiple mysticete calls is described.  
58 The method has been designed to meet the following requirements: (i) a rejection option is im-  
59 plemented, (ii) the classifier is modular, (iii) it is tuned by a very few (easy-to-set) parameters  
60 and (iv) it involves a compression option so as to provide a good trade-off between robustness to  
61 call variability and computational load. The proposed approach relies on the sparse framework  
62 recently developed in signal processing and machine learning [34–36]. Sparse representations ex-  
63 press a given signal as a linear combination of base elements in which many of the coefficients are

64 zero. Such representations can capture the possible variability observed for some vocalizations  
65 and can automatically be learned from the time-series of the digitized acoustic signals, without  
66 requiring prior transforms such as spectrograms, wavelets or cepstrums. This framework is gen-  
67 eral and applicable to any mysticete call lying in a linear subspace described by a dictionary-based  
68 representation. Successfully applied to the detection of mysticete calls [23], this framework is thus  
69 extended to the classification of mysticete calls and evaluated in this context. To the authors' best  
70 knowledge, this paper is a first attempt in this direction.

71 The paper is organized as follows. In Sec. II, the classification method is presented. The  
72 performance of the classifier is then evaluated on five call classes extracted from four real datasets  
73 in Sec. III. Finally, conclusions are given in Sec. IV.

74 **Notation:** Throughout this paper,  $\mathbb{R}^n$  designates the space of all  $n$ -dimensional real column  
75 vectors and  $\mathbb{R}^{n \times m}$  is the set of all real matrices with  $n$  rows and  $m$  columns. The superscript  $T$   
76 means transposition.  $\|\cdot\|_p$  designates the  $\ell_p$  norm.

## 77 II. METHODOLOGY

78 Supervised learning makes it possible for systems to perform automatic classification of pre-  
79 viously unseen inputs, after learning examples labeled by experts. The learning phase proceeds  
80 as follows. A labeled or training dataset is made of  $N$  pairs  $\{(\mathbf{s}_i, \ell_i)\}_{1 \leq i \leq N}$  representative of  $C$   
81 classes, i.e.,  $C$  call types in our case, where  $\mathbf{s}_i$  is the  $i$ -th feature vector in the training set and  $\ell_i$   
82 is the corresponding class or label of  $\mathbf{s}_i$ , e.g.,  $\ell_5 = 3$  means that the fifth element of the *training*  
83 *set* belongs to the third class. This training set is used to determine a map  $f(\cdot | \{(\mathbf{s}_i, \ell_i)\}_{1 \leq i \leq N})$  that  
84 infers a label from a given feature vector.

85 The map  $f$  is either learned on the training set by minimizing a loss function representing the  
86 cost paid for inaccuracy of predictions (i.e., discrepancy between the predicted and the actual la-  
87 bel) or derived from a prior choice of a *similarity measure* that compares new test data to training  
88 examples. Neural network-based classifiers typically implement the first approach, whereas meth-  
89 ods such as banks of matched-filters [37] or spectrogram correlators [32, 38] implement the second  
90 one.

91 As discussed below, our method relies on the second approach. This choice is mainly motivated  
92 by the will to build a robust and modular method where the similarity measure does not depend  
93 on the training set or on the number of call classes. It is also desirable to avoid using too many

94 ("no-so-easy-to-tune") hyperparameters so as to ease the deployment of the method.

95 In the sequel,  $\{\mathbf{s}_k : k > N\}$  stands for the *test* feature vectors that the system must classify.  
96 Given such a test feature vector  $\mathbf{s}_k$  with  $k > N$ ,  $\hat{\ell}_k = f(\mathbf{s}_k | \{(\mathbf{s}_i, \ell_i)\}_{1 \leq i \leq N})$  is the output label in  
97  $\{1, 2, \dots, C\}$  assigned to  $\mathbf{s}_k$ .

98 In the method proposed below, feature vectors are digitized time-series of calls. It is assumed  
99 that detection of regions of interest within the time-series has already been achieved either au-  
100 tomatically or manually. In Sections II A and II B, the sparse representation and classification  
101 framework for calls is presented. Sections II C and II D introduce the compression and the rejec-  
102 tion options. In Section II E, an overall description of the procedure is given.

### 103 A. From standard similarity measures to sparse representation

104 There exists a wide variety of similarity measures, *e.g.* Euclidean distance, absolute value, like-  
105 lihood, correlation, etc. For instance, let  $|\langle \mathbf{s}_k, \mathbf{s}_i \rangle|$  be the non negative normalized scalar product  
106 or *correlation* between a signal  $\mathbf{s}_k$  and a signal  $\mathbf{s}_i$ . For approaches such as banks of matched filters  
107 or spectrogram correlators, the map  $f$  chooses the class that maximizes the correlation between a  
108 test signal  $\mathbf{s}_k$ ,  $k > N$ , and all the signals in the training dataset, *i.e.*,

$$\hat{\ell}_k = \ell_{i^*}, \quad (1)$$

109 where  $i^* = \operatorname{argmax}_{i \in \{0, 1, \dots, N-1\}} |\langle \mathbf{s}_k, \mathbf{s}_i \rangle|$ .

110 A well-known extension of such an approach is the  $K$  Nearest Neighbors algorithm (KNN)  
111 [39] where  $\mathbf{s}_k$  is assigned to the most common class among its  $K$  nearest neighbors (*e.g.*, the  $K$   
112 signals in the training dataset having the highest correlation with  $\mathbf{s}_k$ ). In general, choosing  $K$   
113 greater than one is beneficial as it reduces the overall noise [40].

114 Beyond KNN, the classification can be based on a similarity measure between the test signal  
115  $\mathbf{s}_k$  to be labeled and a *linear combination* of the  $K$  signals closest to  $\mathbf{s}_k$ . All training signals then  
116 become elementary *atoms* which can be combined to create new signals. In this way, the new  
117 representation space makes it possible to cover a larger space than the original training dataset  
118 and, as such, is expected to better capture the intrinsic/proper structure of the signals of interest.  
119 On the one hand,  $K$  should be small enough to prevent overfitting, especially in presence of noise.  
120 On the other hand, given a test signal, the similarity measure must help select a linear combination

121 of atoms from the same class as the signal to guarantee a meaningful comparison between this one  
 122 and each average model of each class. Therefore, the choice of  $K$  results from a trade-off between  
 123 the risk of overfitting and the necessity to approximate sufficiently well the test signal.

124 Formally, it is assumed that any test signal  $\mathbf{s}_k$  with dimension  $n$  from class  $c$  approximately lies  
 125 in the linear span of the training signals associated with this class, i.e.,

$$\mathbf{s}_k \approx \mathbf{A}_c \mathbf{w}_c, \text{ with } \|\mathbf{w}_c\|_0 \leq K \ll N_c, \quad (2)$$

126 where  $\mathbf{A}_c \in \mathbb{R}^{n \times N_c}$  is a matrix containing all the  $N_c$  training signals of length  $n$  belonging to the  
 127 class  $c$ ,  $\mathbf{w}_c \in \mathbb{R}^{N_c}$  is a vector of weights used in the linear combination and  $\|\mathbf{w}_c\|_0$  denotes the  $\ell_0$ -  
 128 pseudonorm that returns the number of non-zero coefficients in  $\mathbf{w}_c$ . When  $\mathbf{s}_k$  can be represented  
 129 by a small number of non-zero coefficients in the basis  $\mathbf{A}_c$ , model (2) is referred to as “sparse  
 130 representation” in the signal processing literature [35]. The inequality  $\|\mathbf{w}_c\|_0 \leq K$  is called the  
 131 sparsity constraint. This constraint  $K$  is directly related to the “complexity” of each single call to  
 132 be classified. Signals combining variability and high complexity (such as erratic signals) must be  
 133 constructed from a large number of atoms while signals of low complexity should be composed of  
 134 a few atoms. For instance, D calls of blue whales [41] are frequency-modulated (FM) sweep that  
 135 could well be approximated by a linear combination of a few atoms. However, such calls exhibit  
 136 variability in initial frequency, FM rate, duration, and bandwidth. Therefore, the  $\ell_0$  norm of  $\mathbf{w}_c$   
 137 is small for each single call but the active atoms, corresponding to non-zero entries of  $\mathbf{w}_c$ , can be  
 138 different from one call to another so that  $N_c$  must be large. Note that model (2) is an approximation  
 139 as calls may be affected by local propagation conditions and noise. However, the very good results  
 140 obtained in Sec. III indicate that it is sufficiently accurate for classification purposes. Examples of  
 141 test signal reconstruction with training signals are shown in the appendix for real calls.

## 142 B. Sparse Representation-based Classification

143 Based on a linear model similar to (2), Wright et al. proposed a Sparse Representation-based  
 144 Classifier (SRC) in [34]. It achieved impressive results in a wide range of applications such as bird  
 145 classification [42], EEG signal classification [43], face recognition [34, 44]. Originally applied to  
 146 face recognition, we suggest adapting this approach to our context. To this end, this subsection  
 147 recalls the SRC procedure, whereas the next two propose additional features to improve SRC

148 performance in our particular application.

149 SRC assumes that test signals can be represented by a linear combination of training signals.  
 150 In our context, these signals are digitized time-series and represent the input feature vectors of  
 151 the classifier. SRC is a two-step procedure: (i) it seeks the linear combination of training signals  
 152 that best approximates — in the sparse sense — the test signal and (ii) chooses the class that  
 153 mostly contributes to this approximation. More precisely, the true label of the test signal  $\mathbf{s}_k$  being  
 154 unknown,  $\mathbf{s}_k$  is first represented as a linear combination of all training signals stored in a matrix  
 155  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_C] \in \mathbb{R}^{n \times \sum_{c=1}^C N_c}$ , where  $C$  is the number of call classes, i.e.,

$$\mathbf{s}_k \approx \mathbf{A}\mathbf{w}, \text{ with } \|\mathbf{w}\|_0 \leq K. \quad (3)$$

156 Ideally, the entries of  $\mathbf{w} \in \mathbb{R}^{\sum_{c=1}^C N_c}$  are all zeros except at most  $K$  entries related to the training  
 157 signals from the same class as the test signal. For instance, if  $\mathbf{s}_k$  belongs to class  $c$ , i.e.,  $\ell_k = c$ ,  
 158 then  $\mathbf{w}$  should ideally satisfy  $\mathbf{w} = [0, \dots, 0, \mathbf{w}_c^T, 0, \dots, 0]^T$  where  $\mathbf{w}_c \in \mathbb{R}^{N_c}$  and  $\|\mathbf{w}_c\|_0 \leq K$ .  
 159 Therefore, the actual class of the test signal could be obtained by estimating  $\mathbf{w}$  and finding the  
 160 indexes of the nonzero entries of  $\mathbf{w}$ . However, in practice, because of the noise and the non-  
 161 orthogonality between training signals from different classes, nonzero entries of  $\mathbf{w}$  may appear at  
 162 indexes not related to the true class of the test signal. Consequently, the class label for the test  
 163 signal is not determined by finding the indexes of the nonzero entries of  $\mathbf{w}$  but by finding the  
 164 class-specific entries of  $\mathbf{w}$  yielding the best approximation of  $\mathbf{s}_k$  in (3).

165 More specifically, the two-step procedure of SRC is as follows:

- 166 1. Estimate  $\mathbf{w}$  by sparsely encoding  $\mathbf{s}_k$  over the basis  $\mathbf{A}$ . i.e., by solving

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{s}_k - \mathbf{A}\mathbf{w}\|_2^2, \text{ with } \|\mathbf{w}\|_0 \leq K. \quad (4)$$

167 Sparse encoding can be performed with pursuit algorithms [35] or  $\ell_1$ -norm minimization  
 168 [45]. In Section III, this step is implemented with orthogonal matching pursuit (OMP) [46].

- 169 2. Associate  $\mathbf{s}_k$  to the class  $\hat{\ell}_k$  that satisfies

$$\hat{\ell}_k = \underset{1 \leq c \leq C}{\operatorname{argmin}} \|\mathbf{s}_k - \mathbf{A}\delta_c(\mathbf{w}^*)\|_2^2, \quad (5)$$

170 where  $\delta_c(\mathbf{w}^*)$  is a characteristic function that selects the coefficients of  $\mathbf{w}^*$  associated with

171 the  $c$ -th class. For any  $\mathbf{w} \in \mathbb{R}^{\sum_{c=1}^C N_c}$ ,  $\delta_c(\mathbf{w}) \in \mathbb{R}^{\sum_{c=1}^C N_c}$  is a vector whose nonzero  
 172 entries are the entries in  $\mathbf{w}$  that are related to the  $c$ -th class. For instance, if  $\mathbf{w} =$   
 173  $[\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_C^T]^T$  where each  $\mathbf{w}_i$  belongs to class  $i$ , then  $\delta_c(\mathbf{w}) = [0, \dots, 0, \mathbf{w}_c^T, 0, \dots, 0]^T$ .  
 174 The solution to (5) is found by exhaustive search through all the classes.

### 175 C. Compression option

176 Ideally, the training dataset  $\mathbf{A}$  should span the space that includes any mysticete call we wish  
 177 to classify. In particular, for each class,  $\mathbf{A}_c$  should incorporate enough variability to model all  
 178 possible calls of the same class. It is thus desirable to inject in  $\mathbf{A}$  the maximum amount of infor-  
 179 mation we have on these calls. However, the computational complexity of (4) grows with the size  
 180 of  $\mathbf{A}$  without necessarily adding any performance improvement if  $\mathbf{A}$  contains redundant signals.  
 181 To limit redundancy in  $\mathbf{A}$  and thus achieve a trade-off between variability and computational load,  
 182 we suggest building a lower dimensional dictionary  $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_C]$  from the training  
 183 dataset, where each submatrix  $\mathbf{D}_c$  has  $N'_c \leq N_c$  columns, i.e.,  $\mathbf{D}_c \in \mathbb{R}^{n \times N'_c}$ . Each  $\mathbf{D}_c$  is found as  
 184 the subdictionary that leads to the best possible representation for each training signal of class  $c$   
 185 with the sparsity constraint (4). More precisely, the new subdictionary  $\mathbf{D}_c$  for class  $c$  is derived by  
 186 solving the minimization problem:

$$\begin{aligned} & \min_{\mathbf{D}_c, \mathbf{W}} \|\mathbf{A}_c - \mathbf{D}_c \mathbf{W}\|_F^2 \\ & \text{subject to } \|\mathbf{w}_i\|_0 \leq K, \forall 1 \leq i \leq N_c, \end{aligned} \quad (6)$$

187 where  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{N_c}]$  and  $\mathbf{w}_i \in \mathbb{R}^{N'_c}$ . The minimization problem (6) is commonly referred  
 188 to as “dictionary learning” and is only performed offline once. Numerical solutions to (6) can  
 189 be obtained with the method of optimized direction (MOD) [47], K-SVD [48] or online learning  
 190 [45]. Once the lower dimensional dictionary is learned,  $\mathbf{A}$  and  $\mathbf{A}_c$  are replaced by  $\mathbf{D}$  and  $\mathbf{D}_c$  in (4)  
 191 and (5), respectively and  $\delta_c(\cdot)$  is adapted to the size of  $\mathbf{D}$ . In addition to removing the redundant  
 192 information in the learning process, dictionary learning extracts the salient feature of  $\mathbf{A}$  and this  
 193 thus expected to limit the sensitivity to noisy training signals or to overfitting issues.

194 **D. Rejection option**

195 A major challenge in automatic classification of underwater sounds is the management of  
 196 “noise”. In our context, noise is defined as any test signal, fed into the classifier, that does not  
 197 belong to one of the  $C$  output mysticete call classes of the classifier. This noise can be:

- 198 • Transient noise or interference that designates any transient signal of no interest for the  
 199 classifier, *e.g.* calls of other whales, ship noise, airguns, earthquakes, ice tremors, etc.
- 200 • Background noise which is a mixture of numerous unidentifiable ambient sound sources  
 201 that does not include any transient signal.

202 The rejection option offers the capability of refusing to assign the examined signal to any class,  
 203 possibly prompting for a deeper investigation by a human analyst. In [34, Sec. 2.4], a rejection  
 204 option is proposed for SRC. It relies on the assumption that a valid test signal has a sparse repre-  
 205 sentation whose nonzero entries concentrate mostly on one class, whereas a signal to be rejected  
 206 has coefficients spread widely among multiple classes. While such an assumption may be valid  
 207 in applications such as face recognition [34], it is not applicable in our context. The main reason  
 208 is that transient underwater acoustic noises may have a non-negligible amount of their energy ly-  
 209 ing in a subspace in which a specific class of calls resides. For instance, the sparse coefficients  
 210 of impulsive noise are likely to concentrate on classes related to impulsive calls (such as the fin  
 211 whale 20 Hz calls presented in Sec. III A), whereas tonal noise coefficients will be related to tonal  
 212 calls having similar frequencies. To deal with noise, we propose to apply a post-processing pro-  
 213 cedure that decides whether the test signal actually lies in the subspace spanned by the column of  
 214 the subdictionary corresponding to the class chosen by SRC. More precisely, the result of SRC is  
 215 validated if the estimated Signal-to-Interference-plus-Noise Ratio (SINR)

$$\text{SINR}(\mathbf{s}_k, \hat{\ell}_k) = \frac{\|\mathbf{D}\delta_{\hat{\ell}_k}(\mathbf{w}^*)\|_2^2}{\|\mathbf{s}_k - \mathbf{D}\delta_{\hat{\ell}_k}(\mathbf{w}^*)\|_2^2} \quad (7)$$

216 is greater than some threshold. Based on model (2),  $\mathbf{D}\delta_{\hat{\ell}_k}(\mathbf{w}^*)$  is an estimate of the signal of in-  
 217 terest and  $\mathbf{s}_k - \mathbf{D}\delta_{\hat{\ell}_k}(\mathbf{w}^*)$  is an estimate of the interference plus background noise. This criterion  
 218 measures the reconstruction quality of the test signal  $\mathbf{s}_k$  when approximated by a linear combi-  
 219 nation of the elements of  $\mathbf{D}_{\hat{\ell}_k}$ . It is inspired by Constant False Alarm Rate (CFAR) detectors of  
 220 known signal in noise with unknown power, which show optimal properties with respect to de-  
 221 tection performance [22, 23, 49]. The methodology used to set the SINR threshold is presented

222 in Sec. III C 2. A key aspect of our approach is that the classifier does not need to learn features  
 223 of transient noises to reject them. This differs from methods such as [25] where noise features  
 224 are learned by neural networks or from [24] where, for each class of noise, “*a parametric model*  
 225 *of noise is introduced. The models are based on the spectral properties of typical kinds of im-*  
 226 *pulsive noise observed in the data*” [24, pp. 360]. This implies to find exhaustive examples of  
 227 underwater noise, which seems difficult given the complexity of the underwater environment. The  
 228 characteristics of sensed underwater sounds are highly dependent on the anthropogenic, biological,  
 229 geological or oceanographic environment as well as on the way sensors are mounted in the  
 230 water column. So the noise learned or modeled in one context can hardly be transposed to another  
 231 one.

### 232 E. Overall procedure

233 The classification process resulting from the foregoing considerations is hereafter referred to  
 234 as SINR-SRC. It is summarized as follows and illustrated with two classes in Figure 1.

- 235 1. Offline selection of training signals representative of their call class.
- 236 2. Offline application of the compression option (6) if required.
3. Given some test signal  $\mathbf{s}_k$ , perform a sparse encoding of  $\mathbf{s}_k$  over dictionary  $\mathbf{D}$  by computing:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{s}_k - \mathbf{D}\mathbf{w}\|_2^2, \text{ with } \|\mathbf{w}\|_0 \leq K.$$

4. Application of SRC by computing the class contributing most to the test signal  $\mathbf{s}_k$ :

$$\hat{\ell}_k = \underset{1 \leq c \leq C}{\operatorname{argmin}} \|\mathbf{s}_k - \mathbf{D}_c \delta_c(\mathbf{w}^*)\|_2^2.$$

- 237 5. Application of the rejection option: if  $\operatorname{SINR}(\mathbf{s}_k, \hat{\ell}_k)$  is greater than some threshold, the result  
 238 provided by SRC is validated, otherwise  $\mathbf{s}_k$  is considered as noise.

239 This SINR-SRC procedure can be illustrated by the scheme shown in Fig. 1.

240 In addition to the good classification performance achieved by SINR-SRC (see Sec. III), note  
 241 also that it is modular, which can be very useful in an operational context. For instance, if a new

242 class of mysticete calls must be added to an existing SINR-SRC classifier, there is no need to  
 243 “retrain” the entire classifier as required in approaches such as neural networks, random forest or  
 244 support vector machine. Only the new subdictionary associated to the new class must be learned.  
 245 Moreover, to reduce miss-classifications of online passive acoustic monitoring, prior information  
 246 such as the geographical position of the sensor could be taken into account by removing the sub-  
 247 dictionaries in  $D$  corresponding to species whose habitats are known to be far away from the  
 248 sensor.

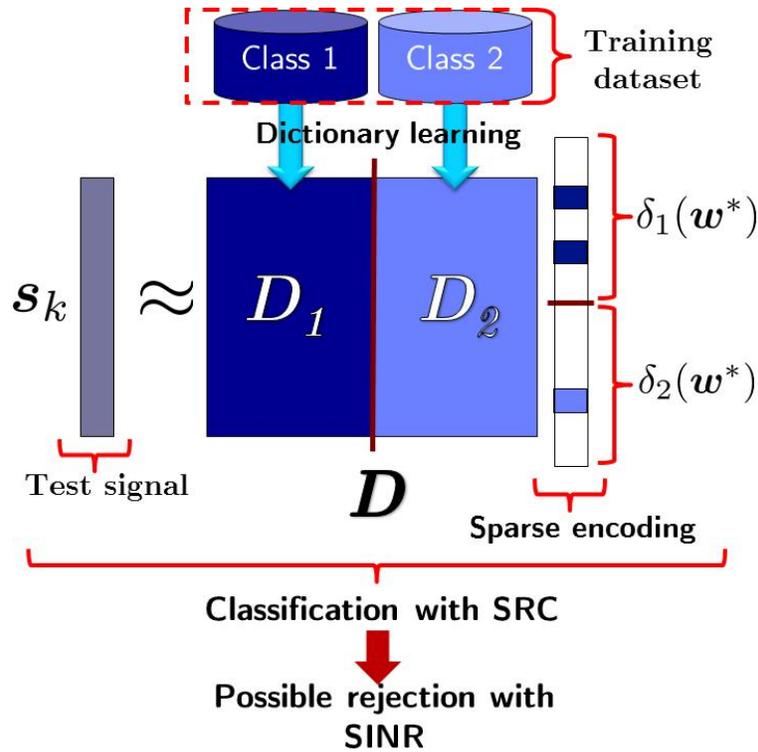


Figure 1. Overview of the classification method for 2 classes.

### 249 III. EXPERIMENTAL RESULTS

#### 250 A. Call library

251 SINR-SRC is evaluated for five call types: Antarctic blue whale Z-calls [50, 51], two types of  
 252 Madagascar pygmy blue whale calls [50], fin whale 20 Hz calls [52], North-Pacific blue whale  
 253 D-calls [26, 41]. These calls have been chosen because:

- 254 • They all overlap in frequencies and some of them have similar durations so they cannot be

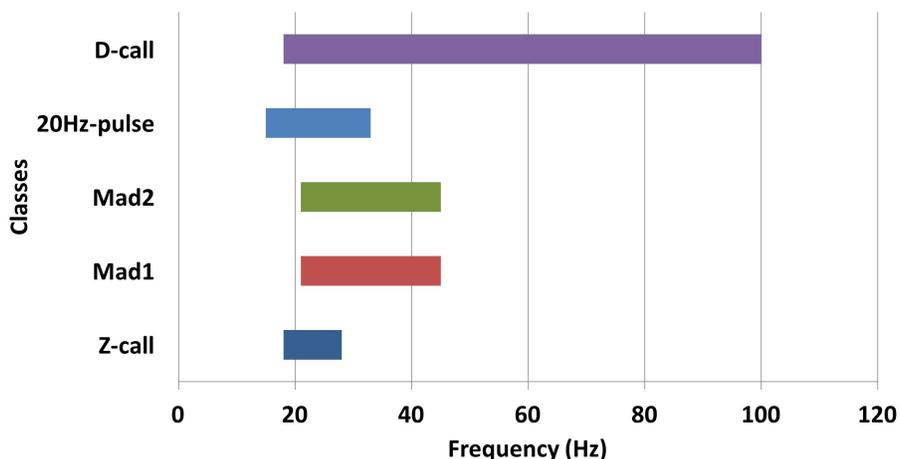


Figure 2. Frequency range of each call type.

255 discriminated based on these two elementary features (Fig. 2 and 3).

- 256 • They offer some variety in terms of signal types: pulsive, tonal sounds or frequency-  
257 modulated (FM) sweeps (Fig. 4).
- 258 • They exhibit different levels of variability: from almost stereotyped (*e.g.*, Z-calls) to variable  
259 in duration, bandwidth and FM rate (*e.g.*, D-calls).

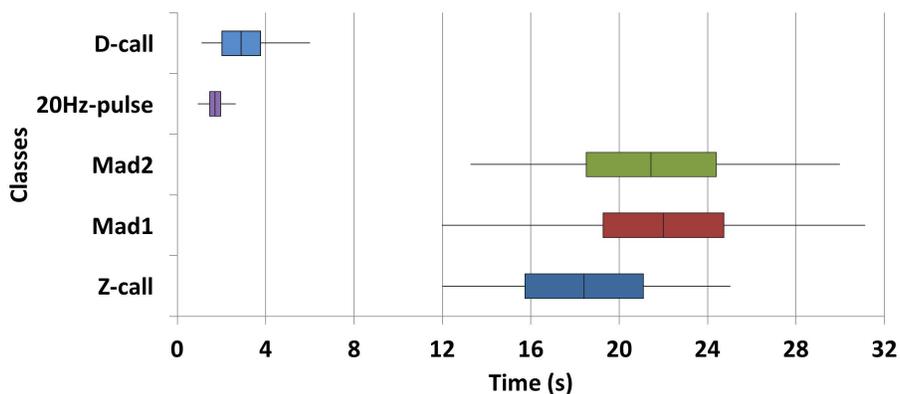


Figure 3. Boxplot of durations for each call type.

260 The five call types were manually extracted from three datasets.

261 *The DEFLOHYDRO dataset:* Three autonomous hydrophones were deployed near the French  
262 territories in the Southern Indian Ocean from October 2006 to January and April 2008. The ob-  
263 jective of the project was to monitor low-frequency acoustic signals, including those produced  
264 by large whales [53]. The three instruments were widely spaced and located in the Madagascar

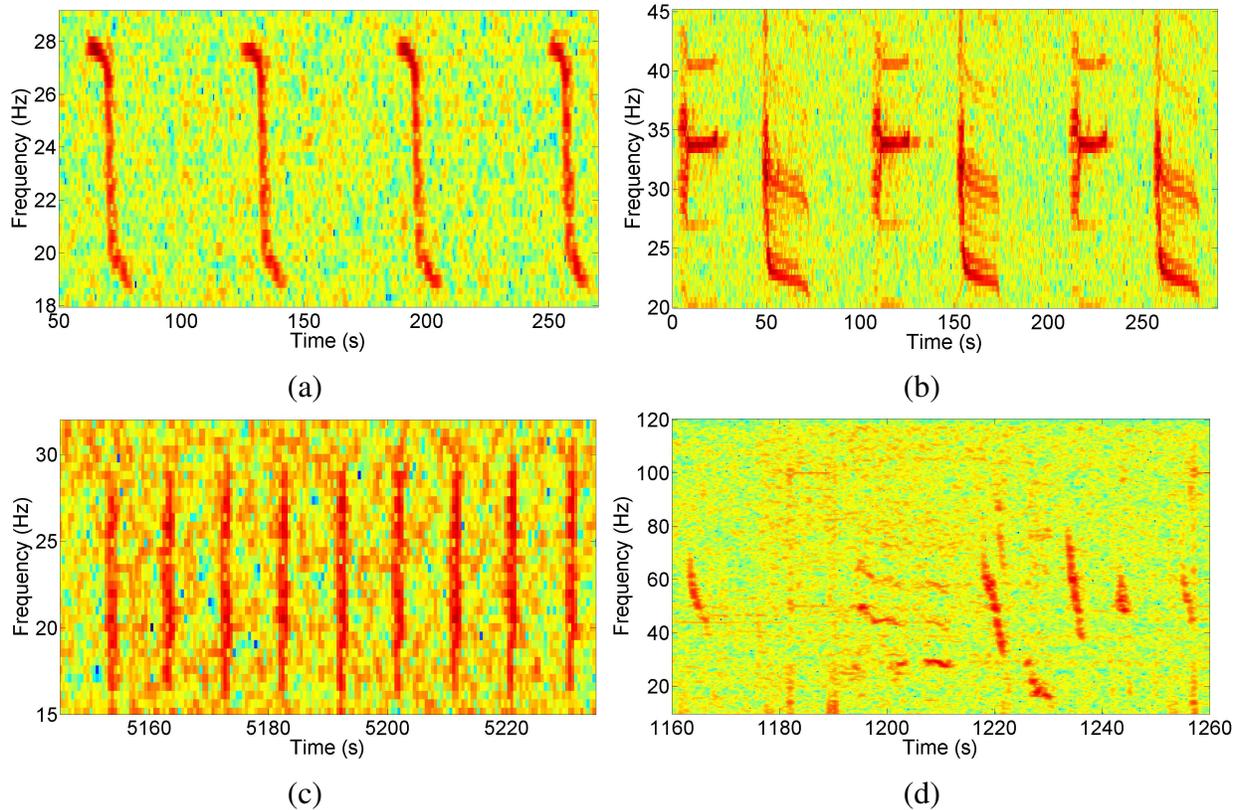


Figure 4. Examples of spectrograms from the call library. (a) four Z-calls produced by Antarctic blue whales, (b) two types of alternative calls produced by Madagascar pygmy blue whales, (c) 20 Hz pulse train produced by fin whales, (d) five D-calls produced by North-Pacific blue whales.

265 Basin, about 320 nautical miles (nm) south of La Reunion Island, and 470 nm to the northeast  
 266 (NEAMS) and 350 nm to the southwest (SWAMS) of Amsterdam Island. The mooring lines were  
 267 anchored on the seafloor between 3410 and 5220 m depths and the hydrophones were deployed  
 268 near the sound channel axis (SOFAR) between 1000 m and 1300 m. The instruments recorded  
 269 sounds continuously at a sampling rate of 250 Hz (frequency range 0.1-110 Hz) [50]. 254 Z-calls  
 270 and 1000 fin whale 20 Hz calls were manually extracted from this dataset.

271 *The OHASISBIO dataset:* In continuation to the DEFLOHYDRO experiment, a network of  
 272 hydrophones was initially deployed in December 2009 at five sites in the Southern Indian Ocean.  
 273 This experiment was designed to monitor low-frequency sounds, produced by seismic and vol-  
 274 canic events, and by large baleen whales [17, 54]. 551 Madagascar pygmy blue whale calls were  
 275 manually extracted from the data recorded by La Reunion Island hydrophone in the Madagascar  
 276 Basin (geographic coordinates : +26° 05' S, +058 °08' E) in May 2015. 264 were type-1 calls and  
 277 287 were type-2, see Fig. 4.

278 *The DCLDE 2015 dataset:* These data have been obtained with high-frequency acoustic record-  
 279 ing packages deployed in the Southern California Bight. 380 D-calls were extracted from data  
 280 recorded at the CINMS B site (latitude: +34° 17' N, longitude: +120° 01' 7" W) in summer 2012  
 281 [26].

282

283 The whole library is composed of 2185 mysticete calls. Each call has been manually annotated  
 284 in time and frequency: start and end time are identified as well as lowest and highest frequency of  
 285 each call. All calls are band-pass filtered according to their annotation and resampled at 250 Hz.  
 286 To apply SRC, all calls must have the same number of time samples, which is easily achieved by  
 287 zero-padding. As shown in Fig. 5, the library contains signals with a large variety of Signal-to-  
 288 Noise Ratios (SNR). The SNR is here defined as the ratio of signal power to noise power, *measured*  
 289 *in the frequency band of each individual call.*

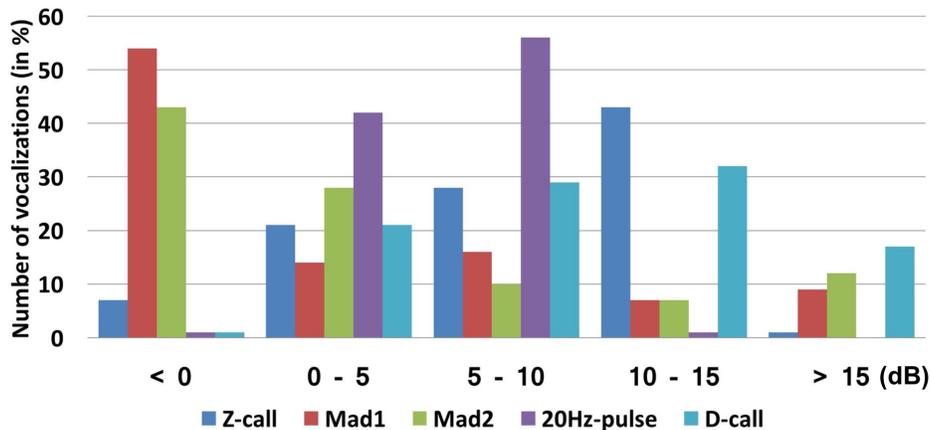


Figure 5. Distributions of the SNRs (in dB) of all the vocalizations in the dataset.

290 Note that four types of calls (Z-calls, 20Hz pulses, Mad1, Mad2) were recorded in the Indian  
 291 ocean and one type (D-calls) in the Southern California Bight. Sensors of the OHASISBIO or  
 292 DEFLOHYDRO networks can sense the first four types of calls in the same recordings [55] but  
 293 North-Pacific blue whales D-calls are observed separately. In practice, this type of D-calls can  
 294 therefore be differentiated from the other calls based on the assumed habitats. To challenge our  
 295 method, location information was not taken into account. A similar approach was considered in  
 296 [25]. In addition, blue whales in the Indian ocean also produce D-calls [56]. Although slightly  
 297 different from D-calls of North-Pacific blue whales, these D-calls are also FM-like signals with  
 298 variable initial frequency, FM rate, duration, and bandwidth. This suggests that our method could  
 299 be relevant for these calls as well.

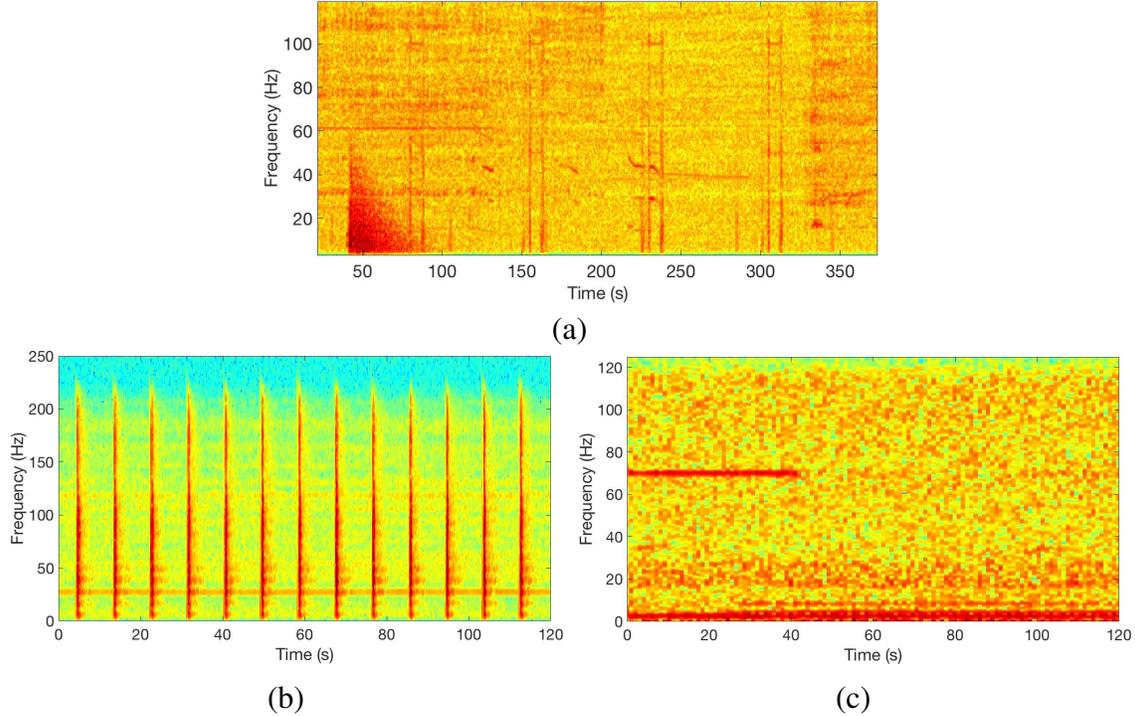


Figure 6. Examples of spectrograms from the noise library. (a) extracted from DCLDE 2015, (b) seismic survey noise provided by Sercel [57] and (c) oceanic noise extracted from DEFLOHYDRO.

## B. Noise library

To test the robustness of SINR-SRC against noise, a noise library was also created. 5000 noise samples were extracted from the DEFLOHYDRO dataset, 5000 from the DCLDE 2015 dataset and 5000 more from a dataset, provided by Sercel [57], recorded during seismic surveys. The first 5000 noise samples mainly correspond to what is called “background noise” in Sec. IID and the others are mostly transient signals of no interest for the classifier, i.e., “interference” (see Fig. 6). In practice, the features (duration, bandwidth, power, etc.) of the noise samples injected into the classifier depends on the actual behavior of the detector used to identify the region of interest before classification. Since we would like to test the performance of our classifier irrespective of the detector, the noise samples were randomly extracted from the datasets. In addition, to challenge the method, noise samples were filtered so that their bandwidths and durations were chosen identical to bandwidths and durations of mysticete calls to be classified. This corresponds to a worst-case scenario for the classifier as filtered noise samples will have a greater amount of energy in the subspaces in which calls reside, leading to an increase of SINR (7).

## 314 C. Performance

315 The performance of SINR-SRC is first analyzed and compared with an implementation of a  
316 state-of-the-art method [29], in the absence of a rejection option. Results with the rejection option  
317 activated are then presented. The impact of the dictionary size as well as the sparsity constraint  
318 is discussed at the end of this section. The performance of the classifier is measured using cross-  
319 validation. As shown in Table I, for each class (with the exception of noise), 100 calls are randomly  
320 selected for training and the remaining calls in this class are used for testing. All the tests presented  
321 are averaged over 100 random selections of the training set to ensure that the results and conclu-  
322 sions do not depend on any specific choice of the training data. For each class, the recall metric,  
323 used below, is defined as the ratio between calls correctly classified and the total number of call  
324 in this class. This metric is sometimes referred to as sensitivity or true positive rate. A recall of  
325 100% for Z-calls class means that all Z-calls have been correctly classified.

### 326 I. Results without rejection

327 Table II shows the average confusion matrix of the SRC algorithm without rejection and without  
328 injecting noise in the classifier. Each column of the matrix represents the percentage of calls in a  
329 predicted class while each row represents the percentage of calls in an actual class. The standard  
330 deviation of the classification results is also displayed in Table II. For this test, no reduction of the  
331 dictionary dimension is applied, i.e.,  $D = A$  and the sparsity constraint  $K$  is set to 3 (impact of  
332 these parameters on the classification performance is discussed in Sec. III C 2). An overall average  
333 recall of 99% is obtained. The SRC classifier not only makes very few errors but is also robust to  
334 training dataset changes.

335 For comparison, Table III displays the classification results obtained with an implementation  
336 of the time-frequency based method introduced in [29]. Similarly to SINR-SRC, this method is  
337 modular and is endowed with a rejection option that requires no noise training. It relies on the  
338 extraction of four amplitude-weighted time-frequency attributes: the average frequency, the fre-  
339 quency variation, the time variation, and the slope of the pitch track in time-frequency space. In  
340 our implementation inspired by [29], this extraction is performed on several spectrograms, each  
341 spectrogram being tuned to the time-frequency features of a specific class. The attributes extracted  
342 from each spectrogram are aggregated and then used as inputs of a quadratic discriminant func-

343 tion analysis classifier. This method yields slightly worse performance than SINR-SRC (without  
344 rejection option). Its average recall is 92.36% compared to 99.46% for SINR-SRC. Note also that  
345 SINR-SRC provides much smaller standard deviations. The method inspired by [29] learns an  
346 average model for each call class and is therefore strongly dependent on the quality of the training  
347 calls. When the training database contains no "outliers", the resulting model is accurate and leads  
348 to good classification results. However, in presence of a few calls with poor quality, the model  
349 is affected and the performance of such a method decreases. In contrast, the dictionary of SINR-  
350 SRC involves sufficiently many atoms so that the reconstruction of the test signal is always good  
351 enough to yield good classification performance.

## 352 2. Results with the rejection option activated

353 We now illustrate the performance of SINR-SRC when the rejection option is activated. We  
354 recall that, as opposed to alternative methods such as [24, 25], rejection of noise is achieved  
355 without learning or modeling noise features, i.e., no dictionary is built from noise data. An input  
356 is rejected by the classifier if the estimated SINR, obtained by computing (7), is lower than some  
357 threshold. This approach is very efficient to discriminate noise data from calls of interest [23].  
358 There exists numerous ways of setting the rejection threshold. For instance, it can be empirically  
359 chosen by the user according to the context and based on his own experience or it can rely on  
360 performance statistics.

361 For instance, we hereafter present a method that is based on the estimation of a false-alarm  
362 probability as commonly done in the Neyman-Pearson framework for binary hypothesis testing.  
363 Assuming that the probability density function (pdf) of the SINR metric is known when noise  
364 samples are injected into the classifier, a rejection threshold guaranteeing a user-specified false-  
365 alarm probability can then be found. However, since the space of all possible underwater transient  
366 noises is very large, it is hardly possible to know precisely this pdf in practice. Therefore, we  
367 resort to an empirical approach and inject into the classifier synthetic random noise samples to  
368 obtain a pdf from which we can set a threshold. This noise is synthetic so as to be as independent  
369 as possible of a specific dataset. In our experiment, we generate independent and identically  
370 distributed samples following the standard Gaussian distribution. Any variance different from 0  
371 could be used, as the SINR metric is scale invariant. The synthetic noise is then obtained by  
372 filtering these samples in time and frequency. The filters have bandwidths and durations identical

373 to bandwidths and durations of mysticete calls to be classified. As explained in Sec. III B, this  
374 corresponds to a worst-case scenario for our method because such a noise will yield a greater SINR  
375 than noise with any other bandwidth and duration. In practice, actual detectors possibly used ahead  
376 of the classifier are unlikely to trigger the classifier with a false alarm signal whose bandwidth  
377 and duration exactly match those of an actual mysticete call. The consideration of worst-case  
378 scenarios is justified by the will to measure achievable classification performance irrespective of  
379 the detector. Rejection thresholds are estimated on each SINR distribution obtained after injecting  
380 Gaussian samples *into each dictionary*. Figure 7 shows an example of a rejection threshold chosen  
381 by setting a false-alarm probability at 1% on the SINR distribution obtained with filtered Gaussian  
382 samples injected into the Z-call dictionary. Note that distributions other than Gaussian may have  
383 been relevant to model noise samples. However, Fig. 7 indicates that the SINR distribution (in red)  
384 of real noises (not necessarily Gaussian) obtained after SRC is close to the distribution obtained  
385 with Gaussian input samples. Once again, the rejection threshold could be selected with alternative  
386 methods. It is beyond the scope of the paper to thoroughly investigate this point; we rather focus  
387 our attention on the general methodology and the classifier structure.

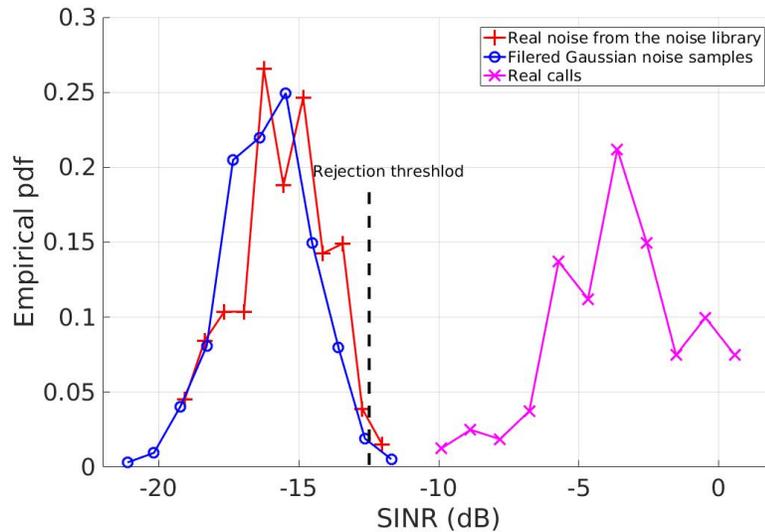


Figure 7. Distribution of SINR, as computed in (7), for Gaussian samples (in blue), real noise (in red) and real calls from the test dataset (in magenta), all identified as Z-calls according to the SRC algorithm without the rejection option. For a 1% false-alarm probability, the rejection threshold is set to -12.5 dB.

388 Table IV shows the average confusion matrix of the SINR-SRC algorithm with rejection. As  
389 expected, activating the rejection option yields a slight drop in the average recall. This drop is  
390 mostly significant for D-calls due to their high variability in duration, frequency range and energy

391 distribution which cause that certain calls in the test dataset are considered as transient noise and  
392 therefore rejected. However, observe that 93.34% of noise inputs are correctly rejected. This  
393 clearly shows that SINR-SRC is capable of efficiently handling input data that are unknown to the  
394 classifier. This property is highly desirable in the low-frequency underwater environment where  
395 interfering sound sources can be very active. The classification results of SINR-SRC with the  
396 rejection option *deactivated* are shown in Table V when noise inputs only are injected into the  
397 classifier. It can be seen that noise inputs are spread among the 5 classes with a slightly higher  
398 probability for classes of calls embedding impulsive structures with a large frequency slope. This  
399 is explained by the large number of transient signals in the noise library.

400 For comparison, the classification results obtained with the method derived from [29], with  
401 its rejection option, are shown in Table VI. A test signal is rejected if the Mahalanobis distance  
402 between its feature vector and its assigned mean attribute vector exceeds 3. This rejection option  
403 does not significantly reduce the recall. However, the noise rejection proposed in [29] is not  
404 as effective as the SINR-SRC rejection option. Actually, Tables IV and VI show that 93.3% of  
405 noise samples are correctly rejected by SINR-SRC, whereas only 66.4% are rejected by [29]. For  
406 a deeper analysis of the rejection performance for SINR-SRC and [29], zooms on the receiver  
407 operating characteristics (ROC) curves are shown in Figures 8 and 9. Such a comparison is all the  
408 more relevant that the noise rejection is controlled by both methods via one parameter only that  
409 we made vary. For our implementation of [29], this parameter is the threshold on the Mahalanobis  
410 distance between a test signal feature vector and its assigned mean attribute. For SINR-SRC,  
411 this parameter is the false alarm probability we can specify to all the SINR distributions obtained  
412 after injections of filtered Gaussian noise samples into the dictionaries. Given a specified false  
413 alarm probability for SINR-SRC, or a specified threshold on the Mahalanobis distance for our  
414 implementation of [29], we calculated the actual false alarm rates and recalls obtained by each  
415 method in presence of real noise and calls. We remind the reader that filtered noise samples have  
416 similar bandwidths and durations as those of mysticete calls to be classified, which is the worst-  
417 case scenario for both methods.

418 These ROC curves highlight the better ability of SINR-SRC to reject noise compared to the  
419 reference method. In particular, the offset in Figure 8 indicates that filtered noise tends to have  
420 average time-frequency attributes close to learned attributes of calls, whatever the type of call.  
421 In the worst-case scenario we have considered, the method derived from [29] cannot provide a  
422 false alarm rate smaller than 5%. Note also the following facts. To begin with, the noise rejection

423 rate of 66.41% reported in the confusion matrix of Table VI corresponds to a false alarm rate of  
 424 33.59%. The reader can then verify that the recall values of Table VI can be retrieved from the  
 425 ROC curves of Figure 8. In the same way, given that a specified false alarm probability of 1% on  
 426 the SINR distributions yielded an actual false alarm rate of 6.66% for SINR-SRC (equivalently,  
 427 a noise rejection rejection rate of 93.34% for this method), the recall values displayed in Table  
 428 IV can be obtained from Figure 9. The ROC curves of Figure 9 also emphasize the relevance  
 429 of setting a false alarm probability of 1%, leading to an actual false alarm rate of 6.66%. This  
 430 choice is seemingly a good trade-off between false alarm rate and recall, even for D-calls. Indeed,  
 431 beyond this false alarm probability, increases in false alarm rates become more important than  
 432 gains in recalls.

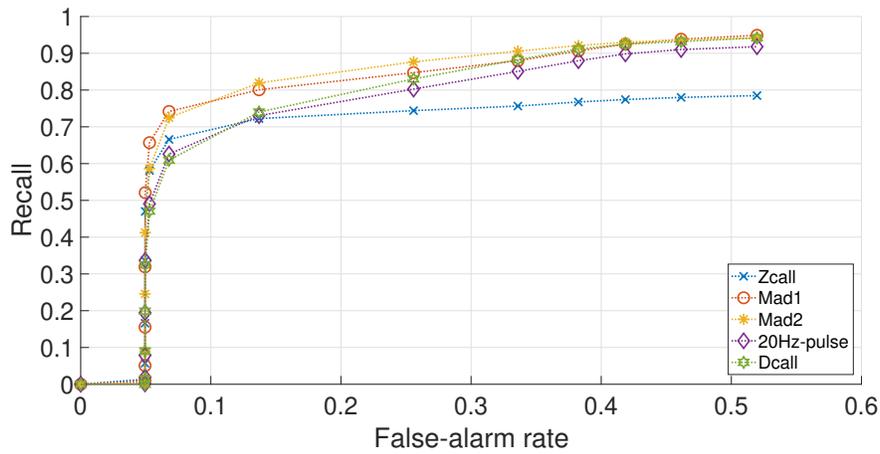


Figure 8. ROC curve for each class of the method derived from [29] with rejection option.

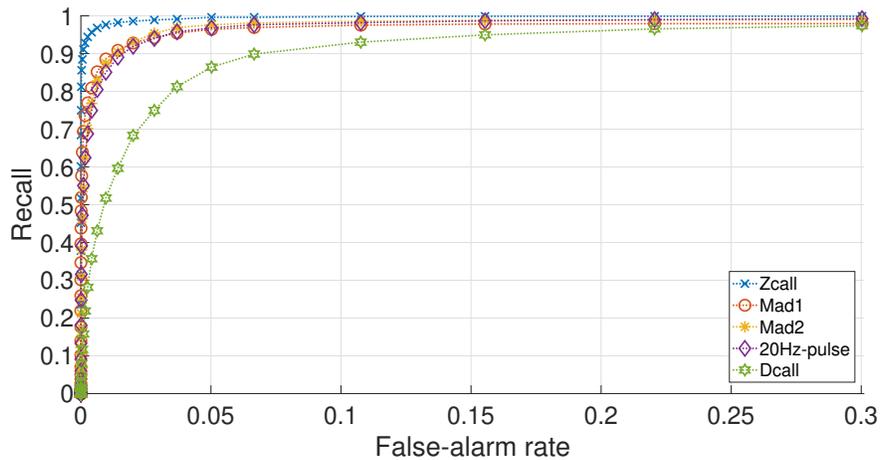


Figure 9. ROC curve for each class of SINR-SRC with rejection option.

433 So far, no reduction of the dictionary dimension has been considered, i.e.,  $D = A$ . As men-

434 tioned in Sec. II B, limiting the redundancy by solving (6) during the training phase may be useful  
 435 to reduce the computational complexity. Figure 10 shows the impact of the dictionary size  $N'_c$   
 436 on the classification performance for each call class. For this test, (6) was solved using online  
 437 dictionary learning [45] (the Matlab code is available at [http://spams-devel.gforge.  
 438 inria.fr/](http://spams-devel.gforge.inria.fr/)). The dictionary size affects the recall and it is interesting to note that its impact  
 439 is class-dependent. For stereotyped calls such as Z-calls, the size of the dictionary can be small  
 440 since the dimension of the signal space is related to the call variability, which is low in this case.  
 441 However, for varying signals such as D-calls, which also have overlapping features with 20 Hz-  
 442 pulses, the classification recall increases (on average) with the dictionary size. In this experiment,  
 443 choosing  $N'_c = 40$  for each class is sufficient to achieve close-to-optimal performance.

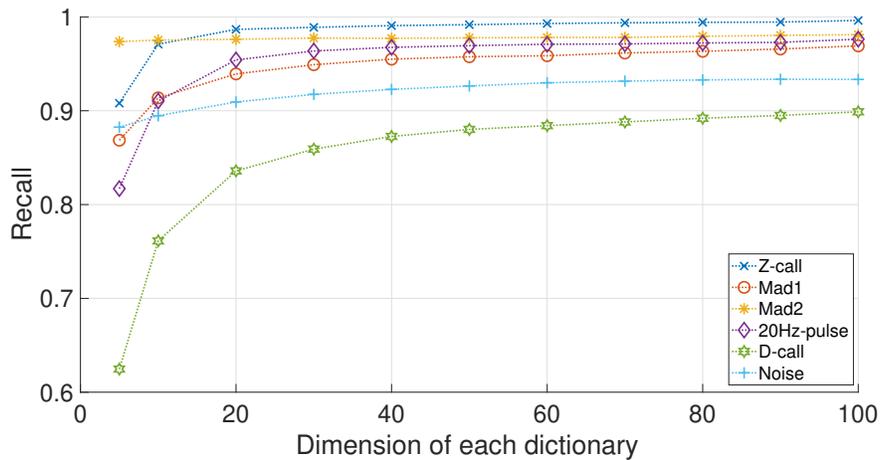


Figure 10. Average recall as a function of the dictionary size  $N'_c$ ,  $K = 3$  and rejection option activated.

444 The impact of the dictionary size on the computational complexity is visible in Figure 11 where  
 445 the run-time-to-signal-duration ratio (RTSDR) of SINR-SRC is shown as a function of the dictio-  
 446 nary size  $N'_c$ . This ratio is computed as the duration of the processing time divided by the total  
 447 duration of the test dataset (58 h). SINR-SRC is implemented in Matlab (without parallel comput-  
 448 ing) and runs on a workstation with the 2.9 GHz Intel Core i7 processor, 8 Gio of RAM memory  
 449 and a DDR3 internal hard drive. Most of the computation time is spent in solving (4) by using  
 450 OMP, which makes the RTSDR increase with  $N'_c$ . In this experiment, the processing time increases  
 451 linearly with  $N'_c$ . Therefore, according to Figure 10, the processing time can be divided by 2.5 by  
 452 choosing  $N'_c = 40$  instead of  $N'_c = 100$  without any performance loss. For  $N'_c = 40$ , SINR-SRC  
 453 took less than 24 seconds to process the 58 hours of tests signals, which meets the requirements  
 454 of most PAM applications. Note that this time is expected to increase with the number of classes

455 considered by the classifier.

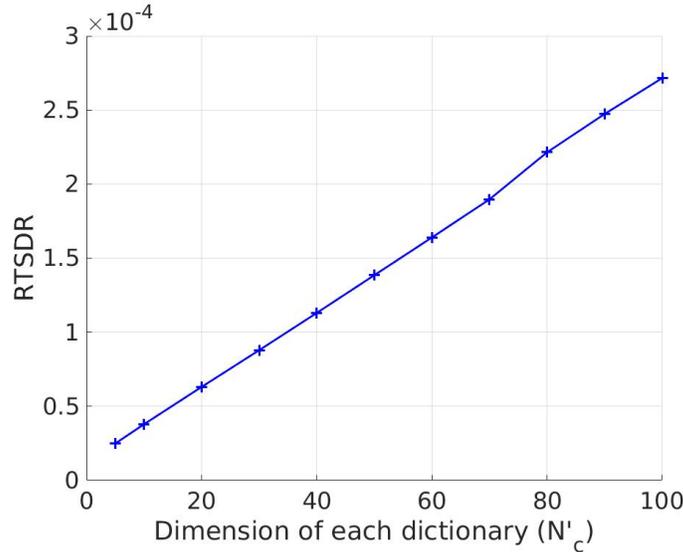


Figure 11. Run-time-to-signal-duration ratio as a function of the dictionary size  $N'_c$ .

456 As shown in Figure 12, the sparsity constraint  $K$  can also affect the classification recall. Sim-  
457 ilarly to the dictionary size, the optimal value for  $K$  depends on the variability and complexity  
458 of the test signals and is therefore class-dependent. However, no fine tuning is required. SINR-  
459 SRC performs better for all classes when  $K$  is greater than 1,  $K = 1$  corresponding to a bank of  
460 matched-filters. For a sparsity constraint greater than 3 and less than 10, this test shows that SINR-  
461 SRC is robust to the choice of  $K$ . Since  $K$  contributes to the complexity of our algorithm, it may  
462 be relevant to limit it to 3 or 4 for the call classes tested in this experiment. In addition, choosing  
463 a large value for  $K$  (much greater than 10 for instance) may be detrimental to the classification  
464 performance as the SINR metric will tend to reject less noise samples [23, Sec. 4.1.2].

#### 465 IV. CONCLUSION

466 Sparse representations have shown to be efficient to classify low frequency mysticete calls.  
467 Such representations model calls as linear combinations of atoms in an (overcomplete) dictionary  
468 in which many of the coefficients are zero. In this framework, the classifier seeks to approximate  
469 the input test signals with (a few) linear combinations of previously learned calls and assigns the  
470 class label that gives the best approximation. The proposed method directly processes the digitized  
471 time series and therefore does not suffer any loss of information due to a possible projection in

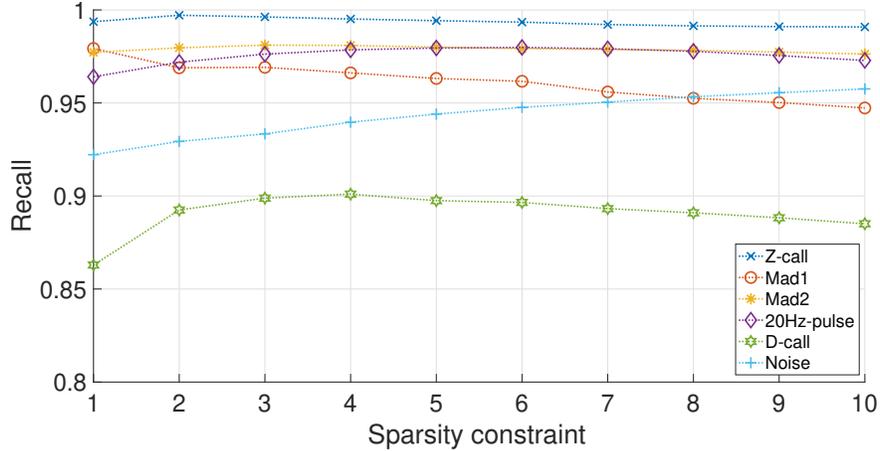


Figure 12. Average recall as a function of the sparsity constraint  $K$ ,  $N'_c = 100$  and rejection option activated.

472 another space (as can be done when extracting features from spectrograms or cepstrums). Since  
 473 the classification is based on a measure of similarity, it relies on a few parameters, namely, the  
 474 dictionary size and the sparsity constraint. These parameters reflect the degree of variability and  
 475 complexity of a given call class. As shown in the numerical experiments, these parameters are  
 476 easy to set and do not require a fine tuning.

477 Sparse representations also allows building simple confidence metrics to reject noise data. The  
 478 SINR statistic (7) has been used at the output of the classifier and has rejected 93.3% of real noise  
 479 data. With this approach, noise is handled without making the algorithm learn the features of real  
 480 noise data. The overall method has been tested on five types of mysticete calls with overlapping  
 481 time-frequency features and different degrees of variability. Numerical results have shown that, on  
 482 the test dataset, 96.4% are correctly classified on average. As expected, stereotyped calls, such as  
 483 Z-calls of Antarctic blue whale are easier to classify than more variable calls such as blue whale  
 484 D calls, which can be incorrectly rejected by the SINR statistic.

485 Class labels can easily be removed or added to the proposed method. This can be useful for  
 486 operational passive acoustic monitoring where prior information such as location of the sensor  
 487 and/or time of the year can be taken into account to focus on specific species.

488 In a recent work [23], sparse representations have shown good performance for detecting mys-  
 489 ticete calls. A possible extension of this work would therefore be to merge both approaches to  
 490 jointly detect and classify mysticete sounds. Since calls are affected by local propagation condi-  
 491 tions and noise, further work could also study the potential benefit of building dictionaries from  
 492 parametric model of calls rather than/as well as from the call themselves. In addition, the SINR

493 statistic could be used as a confidence metric (related to the threshold position) and also as a nov-  
494 elty detector. In this way, the SINR-SRC algorithm would not only offer the capability of rejecting  
495 noise but it could also be used to develop an automatic semi-supervised incremental learning al-  
496 gorithm able to build new dictionaries online. After detection by the SINR-SRC algorithm of an  
497 unknown structured signal, a human analyst could label it and decide to add it to a new dictionary  
498 for automatic classification of future occurrences of this new class of signals.

500 Figures 13 and 14 show examples of Z and D-call reconstruction using Orthogonal Matching  
 501 Pursuit (OMP) [46], with  $K = 3$  atoms. These calls have been extracted from the DEFLOHYDRO  
 502 and the DCLDE 2015 datasets described in Sec. III A.

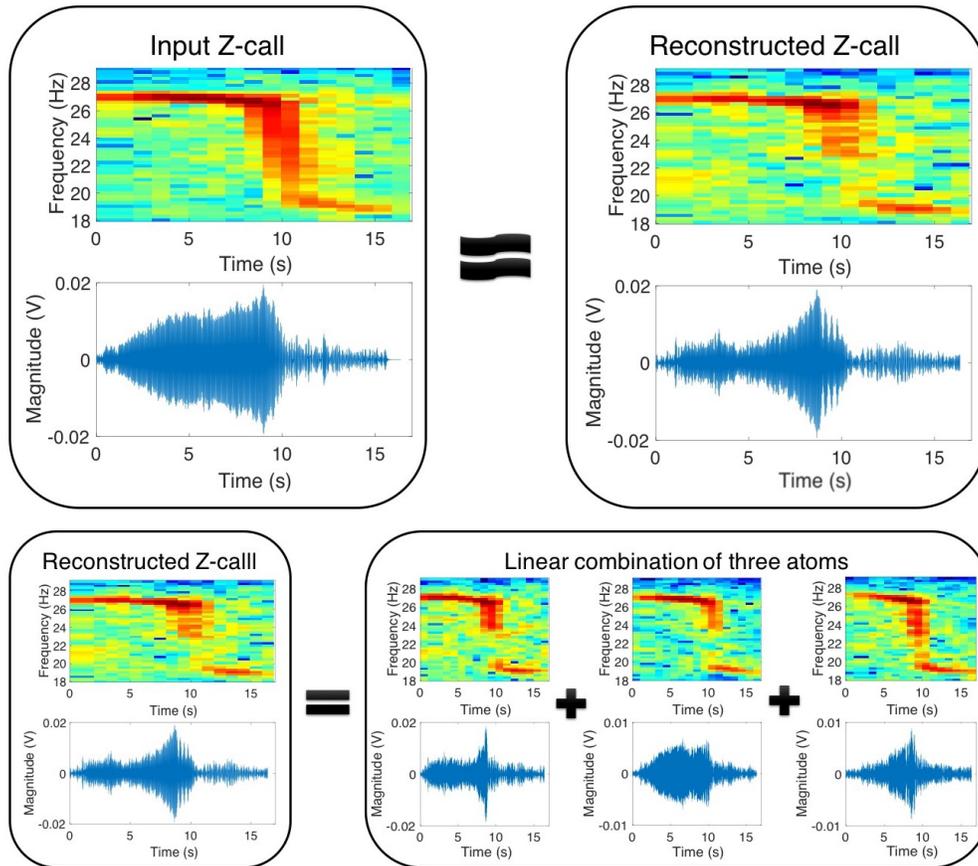


Figure 13. Example of Z-call reconstruction with OMP. The spectrogram representation and the temporal  
 signal of a test Z-call are displayed on the top left. The spectrogram and time representations of the recon-  
 503 structed signal with  $K = 3$  are given on the top right. Below are the three atoms and their combination that  
 504 provided the Z-call reconstruction.

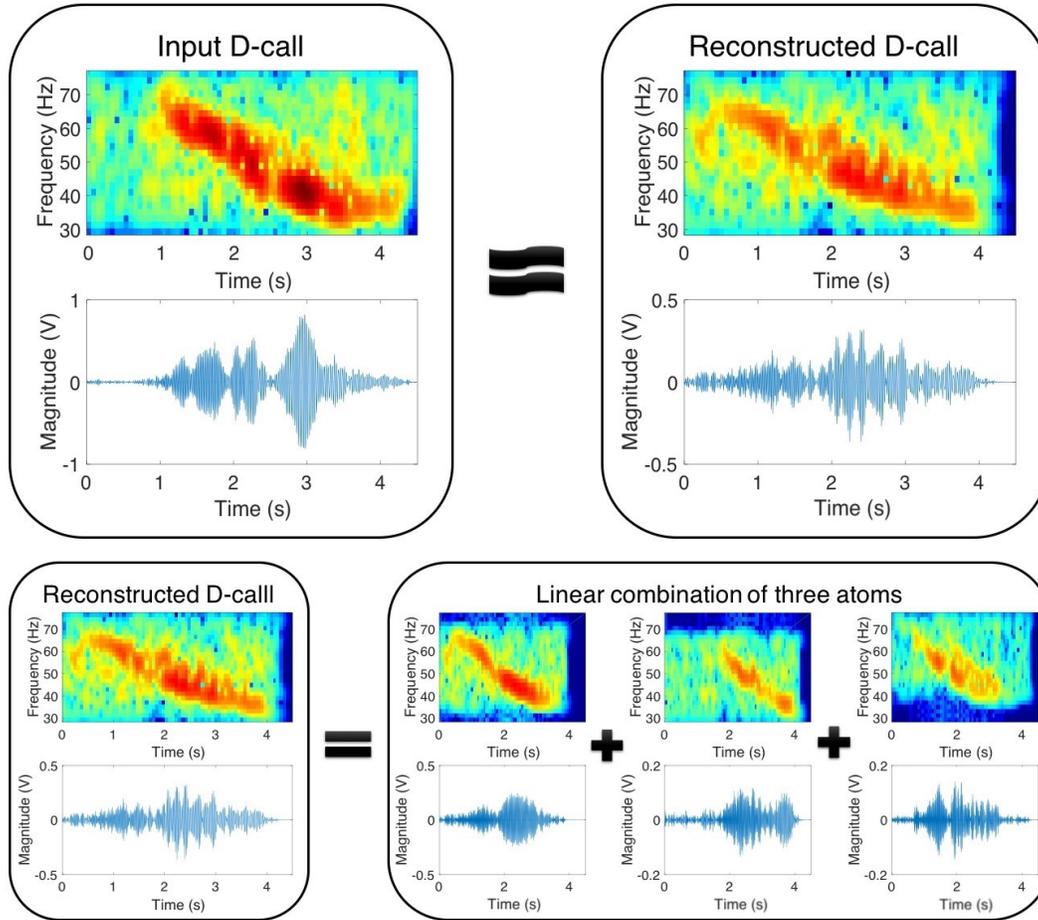


Figure 14. Example of D-call reconstruction with OMP. The spectrogram representation and the temporal signal of a test D-call are displayed on the top left. The spectrogram and time representations of the signal reconstructed by OMP with  $K = 3$  are given on the top right. Below are the three atoms and their combination that provided the D-call reconstruction.

505 **ACKNOWLEDGMENT**

506 This work was funded by Sercel. The authors would like to thank Jean-Yves Royer of the Uni-  
 507 versity of Brest, CNRS Laboratoire Domaines Océaniques for providing the DEFLOHYDRO and  
 508 the OHASISBIO datasets as well as Ana Širović and Simone Baumann-Pickering of the Scripps  
 509 Institution of Oceanography for providing the DCLDE 2015 dataset.

510 The authors are also very thankful to the Associate Editor, Prof. Aaron Thode, whose invaluable

511 comments and remarks made it possible to significantly improve this paper.

---

- 512 [1] D. K. Mellinger, K. M. Stafford, S. E. Moore, R. P. Dziak, and H. Matsumoto, “An overview of fixed  
513 passive acoustic observation methods for cetaceans,” *Oceanography*, vol. 20, December 2007.
- 514 [2] S. E. Parks, C. W. Clark, and P. L. Tyack, “Short- and long-term changes in right whale calling  
515 behavior: The potential effects of noise on acoustic communication,” *J. Acoust. Soc. Am.*, vol. 122,  
516 no. 6, pp. 3725–3731, 2007.
- 517 [3] A. M. Thode, K. H. Kim, S. B. Blackwell, C. R. Greene, C. S. Nation, T. L. McDonald, and A.M  
518 Macrander, “Automated detection and localization of bowhead whale sounds in the presence of seis-  
519 mic airgun surveys,” *J. Acoust. Soc. Am.*, vol. 131, pp. 3726–3747, 2012.
- 520 [4] Renata S. Sousa-Lima, Thomas F. Norris, Julie N. Oswald, and Deborah P. Fernandes, “A review  
521 and Inventory of autonomous recorders fixed autonomous recorders for passive acoustic monitoring  
522 of marine mammals,” *Aquat. Mamm.*, vol. 39, no. 1, pp. 21–28, 2013.
- 523 [5] David K. Mellinger, Stephen W. Martin, Ronald P. Morrissey, Len Thomas, and James J. Yosco, “A  
524 method for detecting whistles, moans, and other frequency contour sounds,” *J. Acoust. Soc. Am.*, vol.  
525 129, no. 6, pp. 4055–4061, 2011.
- 526 [6] Michael Bittle and Alec Duncan, “A review of current marine mammal detection and classification  
527 algorithms for use in automated passive acoustic monitoring,” *Proc. Acoust.*, , no. November, 2013.
- 528 [7] W. M. X. Zimmer, *Passive Acoustic Monitoring of Cetaceans*, Cambridge University Press, 2011.
- 529 [8] M. A. McDonald, J. A. Hildebrand, and S. Mesnick, “Worldwide decline in tonal frequencies of blue  
530 whale songs,” *Endangered Species Research*, vol. 9, no. 1, pp. 13–21, 2009.
- 531 [9] Tzu Hao Lin, Hsin Yi Yu, Chi Fang Chen, and Lien Siang Chou, “Automatic detection and classifica-  
532 tion of cetacean tonal sounds from a long-term marine observatory,” *2013 IEEE Int. Underw. Technol.*  
533 *Symp. UT 2013*, 2013.
- 534 [10] Kathleen M. Stafford, Sue E. Moore, and Christopher G. Fox, “Diel variation in blue whale calls  
535 recorded in the eastern tropical Pacific,” *Anim. Behav.*, vol. 69, no. 4, pp. 951–958, 2005.
- 536 [11] Paul O Thompson, Lloyd T Findley, and Omar Vidal, “20-Hz pulses and other vocalizations of fin  
537 whales, *Balaenoptera physalus*, in the Gulf of California, Mexico,” *J. Acoust. Soc. Am.*, vol. 92, no. 6,  
538 pp. 3051–3057, 1992.
- 539 [12] David K. Mellinger, Carol D. Carson, and Christopher W. Clark, “Characteristics of Minke Whale

- 540 (Balaenoptera Acutorostrata) Pulse Trains Recorded Near Puerto Rico,” *Mar. Mammal Sci.*, vol. 16,  
541 no. 4, pp. 739–756, 2000.
- 542 [13] Hui Ou, W W L Au, and Julie N Oswald, “A non-spectrogram-correlation method of automatically  
543 detecting minke whale boings.,” *J. Acoust. Soc. Am.*, vol. 132, no. 4, pp. EL317—22, 2012.
- 544 [14] Alison K. Stimpert, Whitlow W. L. Au, Susan E. Parks, Thomas Hurst, and David N. Wiley, “Common  
545 humpback whale ( *Megaptera novaeangliae* ) sound types for passive acoustic monitoring,” *J.*  
546 *Acoust. Soc. Am.*, vol. 129, no. 1, pp. 476–482, 2011.
- 547 [15] S. E. Parks, A. Searby, A. Célérier, M. P. Johnson, D. P. Nowacek, and P. L. Tyack, “Sound production  
548 behavior of individual North Atlantic right whales: Implications for passive acoustic monitoring,”  
549 *Endanger. Species Res.*, vol. 15, no. 1, pp. 63–76, 2011.
- 550 [16] Jason Gedamke, Daniel P. Costa, and Andy Dunstan, “Localization and visual verification of a com-  
551 plex minke whale vocalization,” *J. Acoust. Soc. Am.*, vol. 109, no. 6, pp. 3038–3047, 2001.
- 552 [17] E. Leroy, F. Samaran, J. Bonnel, and J.-Y. Royer, “Seasonal and diel vocalization patterns of antarctic  
553 blue whale (*balaenoptera musculus intermedia*) in the southern indian ocean: A multi-year and multi-  
554 site study,” *PloS one*, vol. 11, no. 11, pp. e0163587, 2016.
- 555 [18] Christopher W Clark, Robert Suydam, and Craig George, “Acoustic Monitoring of the Bowhead  
556 Spring Migration off Pt. Barrow, Alaska: Results from 2009 and Status of 2010 Field Effort,” pp. 1–9,  
557 2010.
- 558 [19] Tervo O.M., “Acoustic behaviour of bowhead whales *Balaena mysticetus* in Disko Bay, Western  
559 Greenland,” *Tesis Dr.*, , no. April, pp. 138, 2011.
- 560 [20] Sm Wiggins, Ma McDonald, Lisa M Munger, Sue E Moore, and John Hildebrand, “Waveguide  
561 propagation allows range estimates for North-Pacific right whales in the Bering Sea,” *Can. Acoust.*,  
562 vol. 32, no. 2, pp. 146–154, 2004.
- 563 [21] M. A. Roch, A. Širović, and S. Baumann-Pickering, “Detection, classification, and localization of  
564 cetaceans by groups at the scripps institution of oceanography and san diego state university (2003-  
565 2013),” *Detection, Classification, Localization of Marine Mammals using passive acoustics*, pp. 27–  
566 52, 2013.
- 567 [22] F.-X. Socheleau, E. Leroy, A. Carvallo Pecci, F. Samaran, J. Bonnel, and J.-Y. Royer, “Automated  
568 detection of antarctic blue whale calls,” *J. Acoust. Soc. Am.*, vol. 138, no. 5, pp. 3105–3117, 2015.
- 569 [23] F.-X. Socheleau and F. Samaran, “Detection of mysticete calls: a sparse representation-  
570 based approach,” *IMT Atlantique, Research report RR-2017-04-SC*, Oct. 2017. <https://hal.>

571 [archives-ouvertes.fr/hal-01736178/document](http://archives-ouvertes.fr/hal-01736178/document)

- 572 [24] I. R. Urazghildiiev, C. W. Clark, T. P. Krein, and S. E. Parks, “Detection and recognition of north  
573 atlantic right whale contact calls in the presence of ambient noise,” *IEEE Journal of Oceanic Engi-  
574 neering*, vol. 34, no. 3, pp. 358–368, July 2009.
- 575 [25] X. C. Halkias, S. Paris, and H. Glotin, “Classification of mysticete sounds using machine learning  
576 techniques,” *J. Acoust. Soc. Am.*, vol. 134, no. 5, pp. 3496–3505, 2013.
- 577 [26] “Dclde 2015,” <http://www.cetus.ucsd.edu/dclde/datasetDocumentation.html>,  
578 Accessed: 2016-07-01.
- 579 [27] Carolyn M. Binder and Paul Hines, “Applying automatic aural classification to cetacean vocaliza-  
580 tions,” *Proc. Meet. Acoust.*, vol. 17, pp. 070029, 2012.
- 581 [28] Federica Pace, “Automated classification of humpback whale ( *Megaptera novaeangliae* ) songs using  
582 Hidden Markov Models,” 2013.
- 583 [29] M. F. Baumgartner and S. E. Mussoline, “A generalized baleen whale call detection and classification  
584 system,” *J. Acoust. Soc. Am.*, vol. 139, pp. 2889–2902, 2011.
- 585 [30] X. Mouy, D. Leary, B. Martin, and M. Laurinolli, “A comparison of methods for the automatic  
586 classification of marine mammal vocalizations in the Arctic,” *2008 New Trends Environ. Monit. Using  
587 Passiv. Syst.*, 2008.
- 588 [31] Vasilis Trygonis, Edmund Gerstein, Jim Moir, and Stephen McCulloch, “Vocalization characteristics  
589 of North Atlantic right whale surface active groups in the calving habitat, southeastern United States,”  
590 *J. Acoust. Soc. Am.*, vol. 134, no. 6, pp. 4518–4531, 2013.
- 591 [32] D. K. Mellinger and C. W. Clark, “Recognizing transient low-frequency whale sounds by spectrogram  
592 correlation,” *J. Acoust. Soc. Am.*, vol. 107, pp. 3518–3529, 2000.
- 593 [33] Xavier Mouy, Mohammed Bahoura, and Yvan Simard, “Automatic recognition of fin and blue whale  
594 calls for real-time monitoring in the St. Lawrence,” *J. Acoust. Soc. Am.*, vol. 126, no. 6, pp. 2918–2928,  
595 2009.
- 596 [34] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse  
597 representation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp.  
598 210–227, 2009.
- 599 [35] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image  
600 Processing*, Springer Publishing Company, Incorporated, 1st edition, 2010.
- 601 [36] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*, Cambridge University

- 602 Press, 2012.
- 603 [37] I. R. Urazghildiiev and C. W. Clark, “Acoustic detection of north atlantic right whale contact calls  
604 using the generalized likelihood ratio test,” *J. Acoust. Soc. Am.*, vol. 120, no. 4, pp. 1956–1963, 2006.
- 605 [38] XBAT, “eXtensible BioAcoustic Tool,” [www.birds.cornell.edu/brp/](http://www.birds.cornell.edu/brp/) (date last viewed  
606 14/6/30), Cornell Laboratory of Ornithology, NY, U.S.A.
- 607 [39] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1, Springer series  
608 in statistics Springer, Berlin, 2001.
- 609 [40] T M Cover, and P E Hart, “Nearest Neighbor pattern classification,” *IEEE Trans. Info. Theory*, vol. I,  
610 1967.
- 611 [41] P. O. Thompson, L. T. Findley, O. Vidal, and W. C. Cummings, “Underwater sounds of blue whales,  
612 balaenoptera musculus, in the gulf of california, mexico,” *Marine Mammal Science*, vol. 12, no. 2, pp.  
613 288–293, 1996.
- 614 [42] Lee Ngee Tan, George Kossan, Martin L. Cody, Charles E. Taylor, and Abeer Alwan, “A sparse  
615 representation-based classifier for in-set bird phrase verification and classification with limited training  
616 data,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 763–767, 2013.
- 617 [43] Younghak Shin, Seungchan Lee, Minkyu Ahn, Hohyun Cho, Sung Chan Jun, and Heung No Lee,  
618 “Noise robustness analysis of sparse representation based classification method for non-stationary  
619 EEG signal classification,” *Biomed. Signal Process. Control*, vol. 21, pp. 8–18, 2015.
- 620 [44] Enrique G. Ortiz, Alan Wright, and Mubarak Shah, “Face recognition in movie trailers via mean  
621 sequence sparse representation-based classification,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis.*  
622 *Pattern Recognit.*, pp. 3531–3538, 2013.
- 623 [45] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online learning for matrix factorization and sparse  
624 coding,” *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.
- 625 [46] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function  
626 approximation with applications to wavelet decomposition,” in *Signals, Systems and Computers, 1993.*  
627 *1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, Nov 1993, pp. 40–44 vol.1.
- 628 [47] K. Engan, S. O. Aase, and J. Hakon Husoy, “Method of optimal directions for frame design,” in  
629 *Proceedings of the Acoustics, Speech, and Signal Processing, 1999. On 1999 IEEE International*  
630 *Conference - Volume 05*, Washington, DC, USA, 1999, pp. 2443–2446.
- 631 [48] M. Aharon, M. Elad, and A. Bruckstein, “K-svd: An algorithm for designing overcomplete dictionar-  
632 ies for sparse representation,” *IEEE Trans. Sig. Proc.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

- 633 [49] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-  
634 Wesley, pp. 1 - 524, Reading, Massachusetts, 1991.
- 635 [50] F. Samaran, K. M. Stafford, T. A. Branch, J. Gedamke, J.-Y. Royer, R. P. Dziak, and C. Guinet,  
636 “Seasonal and geographic variation of southern blue whale subspecies in the indian ocean,” *PLoS*  
637 *ONE*, vol. 8, no. 8, pp. 1– 10, 2013.
- 638 [51] K. M. Stafford, D. R. Bohnenstiehl, M. Tolstoy, E. Chapp, D. K. Mellinger, and S. E. Moore, “Antarc-  
639 tic type blue whale calls recorded at low latitudes in the Indian and eastern Pacific Oceans,” *Deep Sea*  
640 *Res., Part I*, vol. 51, pp. 1337–1346, 2004.
- 641 [52] A. Širović, J. A. Hildebrand, S. M. Wiggins, and D. Thiele, “Blue and fin whale acoustic presence  
642 around antarctica during 2003 and 2004,” *Marine Mammal Science*, vol. 25, no. 1, 2009.
- 643 [53] R. P. Dziak, J.-Y. Royer, J. H. Haxel, M. Delatre, and D. R. Bohnenstiehl et al., “Hydroacoustic  
644 detection of recent seafloor volcanic activity in the southern Indian Ocean,” in *Transactions, American*  
645 *Geophysical Union, Fall Meeting, T13. San Francisco, CA (abstract)*, 2008, pp. 1–1.
- 646 [54] E. Tsang-Hin-Sun, J.-Y. Royer, and J. Perrot, “Seismicity and active accretion processes at the  
647 ultraslow-spreading southwest and intermediate-spreading southeast indian ridges from hydroacoustic  
648 data,” *Geophysical Journal International*, vol. 206, no. 2, pp. 1232–1245, 2016.
- 649 [55] E. Leroy, F. Samaran, J. Bonnel J.-Y. Royer, “Identification of two potential whale calls in the southern  
650 Indian Ocean, and their geographic and seasonal occurrence,” *J. Acoust. Soc. Am.*, vol. 142, no. 3, pp.  
651 1413–1427, 2017.
- 652 [56] Rankin S., Ljungblad D., Clark C., and Kato H., “Vocalisations of antarctic blue whales, *balaenoptera*  
653 *musculus intermedia*, recorded during the 2001/2002 and 2002/2003 iwc/sower circumpolar cruises,  
654 area v, Antarctica,” *Journal of Cetacean Research and Management*, vol. 7, pp.13–20, 2005.
- 655 [57] “Sercel,” <http://www.sercel.com/>, Accessed: 2017-03-27.

Class	Training sig.	Test sig.	Total
<b>Z-call</b>	100	154	254
<b>Mad1</b>	100	164	264
<b>Mad2</b>	100	187	287
<b>20Hz-pulse</b>	100	900	1000
<b>D-call</b>	100	280	380
<b>Noise</b>	-	15000	15000

Table I. Number of training and test signals used for each class and for each iteration of the cross-validation.

	Z-call	Mad1	Mad2	20Hz-pulse	D-call
<b>Z-call</b>	<b>100</b>	0.00	0.00	0.00	0.00
	0.10	0.00	0.10	0.00	0.10
<b>Mad1</b>	0.00	<b>97.7</b>	1.90	0.00	0.40
	0.00	1.10	1.10	0.00	0.50
<b>Mad2</b>	0.00	0.30	<b>99.60</b>	0.00	0.10
	0.00	0.30	0.30	0.10	0.20
<b>20Hz-pulse</b>	0.00	0.00	0.00	<b>100</b>	0.00
	0.00	0.00	0.00	0.00	0.00
<b>D-call</b>	0.00	0.00	0.00	0.00	<b>100</b>
	0.00	0.10	0.10	0.00	0.10

Table II. Confusion matrix of the SINR-SRC algorithm (in %) without the rejection option. For each class, the upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials on the call library only.

	<b>Z-call</b>	<b>Mad1</b>	<b>Mad2</b>	<b>20Hz-Pulse</b>	<b>D-call</b>
<b>Z-call</b>	<b>79.89</b>	0.00	19.66	0.45	0.00
	15.96	0.00	16.06	0.44	0.00
<b>Mad1</b>	0.25	<b>96.77</b>	2.70	0.00	0.29
	0.66	1.44	1.22	0.00	0.33
<b>Mad2</b>	3.42	0.69	<b>95.89</b>	0.00	0.00
	3.09	0.37	3.14	0.00	0.00
<b>20Hz-Pulse</b>	0.01	0.00	0.00	<b>93.00</b>	6.99
	0.02	0.00	0.02	5.13	5.14
<b>D-call</b>	3.73	0.00	0.00	0.00	<b>96.27</b>
	1.17	0.00	0.00	0.00	1.17

Table III. Confusion matrix (in %) for the method derived from [29] without rejection option. For each class, the upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials on the call library only.

	<b>Z-call</b>	<b>Mad1</b>	<b>Mad2</b>	<b>20Hz-pulse</b>	<b>D-call</b>	<b>Rejected</b>
<b>Z-call</b>	<b>99.62</b>	0.00	0.00	0.00	0.00	0.38
	0.44	0.00	0.00	0.00	0.00	0.44
<b>Mad1</b>	0.00	<b>96.92</b>	0.52	0.00	0.00	2.56
	0.00	1.27	0.56	0.00	0.00	1.07
<b>Mad2</b>	0.00	0.35	<b>98.11</b>	0.00	0.00	1.54
	0.00	0.28	0.73	0.00	0.00	0.64
<b>20Hz-Pulse</b>	0.00	0.00	0.01	<b>97.63</b>	0.00	2.36
	0.00	0.00	0.03	0.72	0.00	0.72
<b>D-call</b>	0.00	0.01	0.00	0.00	<b>89.89</b>	10.1
	0.00	0.04	0.00	0.00	1.81	1.81
<b>Noise</b>	0.75	0.79	3.21	0.27	1.64	<b>93.34</b>
	0.39	0.62	1.96	0.17	1.89	4.65

Table IV. Confusion matrix of the SINR-SRC algorithm (in %) with the rejection option activated. The false alarm probability specified on the SINR distributions after injection of filtered Gaussian noise samples into the dictionaries is 1%. For each class, the upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials on the call and noise library.

	<b>Z-call</b>	<b>Mad1</b>	<b>Mad2</b>	<b>20Hz-pulse</b>	<b>D-call</b>
<b>Noise</b>	11.76	4.97	35.08	21.70	26.49
	19.61	7.34	27.92	29.49	30.89

Table V. Classification results of SINR-SRC (in %) with noise inputs only. The rejection option is deactivated. The upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials.

	<b>Z-call</b>	<b>Mad1</b>	<b>Mad2</b>	<b>20Hz-Pulse</b>	<b>D-call</b>	<b>Rejected</b>
<b>Z-call</b>	<b>75.64</b>	0.00	0.01	0.00	0.00	24.35
	14.91	0.00	0.06	0.00	0.00	14.91
<b>Mad1</b>	0.01	<b>87.98</b>	1.13	0.00	0.00	10.88
	0.09	3.70	0.61	0.00	0.00	3.55
<b>Mad2</b>	1.93	0.43	<b>90.60</b>	0.00	0.00	7.04
	1.87	0.32	3.88	0.00	0.00	3.26
<b>20Hz-Pulse</b>	0.01	0.00	0.00	<b>85.08</b>	0.00	14.92
	0.02	0.00	0.00	5.30	0.00	5.30
<b>D-call</b>	3.73	0.00	0.00	0.00	<b>88.26</b>	8.01
	1.17	0.00	0.00	0.00	2.51	2.32
<b>Noise</b>	4.94	0.00	21.60	0.00	7.05	<b>66.41</b>
	5.13	0.00	16.68	0.00	5.21	24.62

Table VI. Confusion matrix (in %) for the method derived from [29] with the rejection option activated. The rejection threshold is 3 on the Mahalanobis distance between feature vectors and assigned mean attributes. For each class, the upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials on the call and noise library.

**LIST OF FIGURES**

658	1	Overview of the classification method for 2 classes. ....	11
659	2	Frequency range of each call type. ....	12
660	3	Boxplot of durations for each call type. ....	12
661	4	Examples of spectrograms from the call library. ....	13
662	5	Distributions of the SNRs (in dB) of all the vocalizations in the dataset. ....	14
663	6	Examples of spectrograms from the noise library. (a) extracted from DCLDE	
664		2015, (b) seismic survey noise provided by Sercel [57] and (c) oceanic noise ex-	
665		tracted from DEFLOHYDRO. ....	15
666	7	Threshold estimation on SINR distribution ....	18
667	8	ROC curve for each class of the method derived from [29] with rejection option. ...	20
668	9	ROC curve for each class of SINR-SRC with rejection option. ....	20
669	10	Average recall as a function of the dictionary size $N'_c$ , $K = 3$ and rejection option	
670		activated. ....	21
671	11	Run-time-to-signal-duration ratio as a function of the dictionary size $N'_c$ . ....	22
672	12	Average recall as a function of the sparsity constraint $K$ , $N'_c = 100$ and rejection	
673		option activated. ....	23
674	13	Example of Z-call reconstruction with OMP. The spectrogram representation and	
675		the temporal signal of a test Z-call are displayed on the top left. The spectrogram	
676		and time representations of the reconstructed signal with $K = 3$ are given on the	
677		top right. Below are the three atoms and their combination that provided the Z-call	
678		reconstruction. ....	25
679	14	Example of D-call reconstruction with OMP. The spectrogram representation and	
680		the temporal signal of a test D-call are displayed on the top left. The spectrogram	
681		and time representations of the signal reconstructed by OMP with $K = 3$ are given	
682		on the top right. Below are the three atoms and their combination that provided	
683		the D-call reconstruction. ....	26

684

**LIST OF TABLES**

685

I Number of training and test signals used for each class and for each iteration of the cross-validation. . . . . 32

686

687

II Confusion matrix of the SINR-SRC algorithm (in %) without the rejection option. For each class, the upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials on the call library only. . . . . 32

688

689

690

III Confusion matrix (in %) for the method derived from [29] without rejection option. For each class, the upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials on the call library only. . . . . 33

691

692

693

IV Confusion matrix of the SINR-SRC algorithm (in %) with the rejection option activated. The false alarm probability specified on the SINR distributions after injection of filtered Gaussian noise samples into the dictionaries is 1%. For each class, the upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials on the call and noise library. . . . . 33

694

695

696

697

698

V Classification results of SINR-SRC (in %) with noise inputs only. The rejection option is deactivated. The upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials. . . . . 33

699

700

701

VI Confusion matrix (in %) for the method derived from [29] with the rejection option activated. The rejection threshold is 3 on the Mahalanobis distance between feature vectors and assigned mean attributes. For each class, the upper line contains the mean and the lower line the standard deviation obtained for 100 cross-validation trials on the call and noise library. . . . . 34

702

703

704

705