

Chirplet Transform Applied to Simulated and Real Blue Whale (*Balaenoptera musculus*) Calls

Mohammed Bahoura¹ and Yvan Simard²

¹ Département de Mathématiques, d'Informatique et de Génie
Université du Québec à Rimouski,
300, allée des Ursulines, Rimouski, Québec, Canada, G5L 3A1
Mohammed_Bahoura@uqar.qc.ca

² Institut des Sciences de la Mer,
Université du Québec à Rimouski,
310, allée des Ursulines, Rimouski, Québec, Canada, G5L 3A1
Yvan_Simard@uqar.qc.ca

Abstract. Chirplet transform performance to identify low-frequency blue whale calls is tested with simulations and observations from North-West Atlantic. The three different calls are simulated using linear or quadratic frequency sweeping chirps and a hanning window. The performance of Chirplet transform to accurately estimate the chirp parameters with or without noise is first assessed. Then the performance to classify the real vocalizations from the test dataset using the three features best estimated from the simulations is then assessed. The method has a high classification rate and appears promising to efficiently identify these blue whale signature vocalizations with a reduced number of parameters, even under low signal to noise ratios.

Keywords: Blue whale vocalizations, chirplet transform, feature extraction, vector quantization, noise.

1 Introduction

As birds, marine mammals are highly vocalizing animals and different species can be recognized by their specific sounds, calls or songs. Among them, the blue whale produce powerful low-frequency (< 100 Hz) vocalizations that can propagate over large distances, reaching more than 100 km before being masked by the local ocean noise levels, which varies widely in different environments [1,2]. North Atlantic blue whales produce at least three calls that can be used to identify the species in blind recordings at a given location: the stereotyped A and B infrasonic calls (15-20 Hz) and the more variable audible D or arch call (35-120 Hz) [3,4]. Since about 20 years, such signature calls have been used to track the presence or location of various whale species in different ocean basins along their long-range annual migrations routes [5]. Such passive acoustic monitoring (PAM) of whales requires efficient processing algorithms to detect and identify the signature calls on long recording series, extending to several

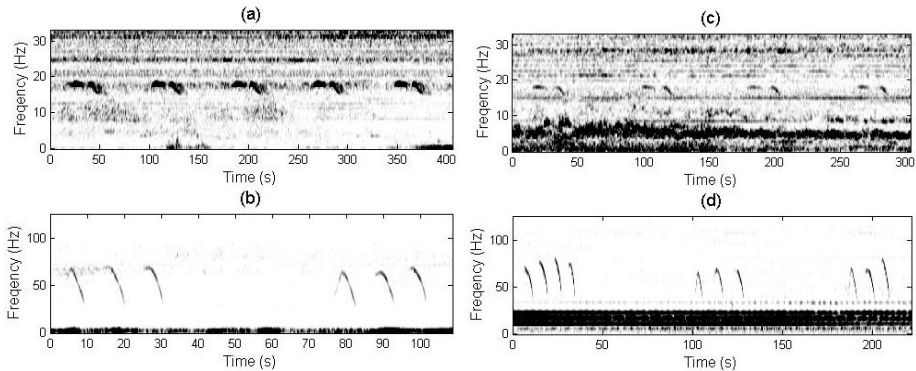


Fig. 1. a) spectrogram of five AB phrases marked with multipath echo, b) spectrogram of six D calls grouped in trio, c) spectrogram of four AB phrases drowned in noise, and d) spectrogram of ten D calls in presence of intense low-frequency noise

months or years Automatic PAM detection algorithms [6] must be efficient, with high detection rate and low false alarms, robust to variable noise conditions and, ideally, fast enough to allow real-time detection and identification under multi-species and multi-sources conditions. Here we present a new method to efficiently identify blue whale signature calls from a limited number of features, which is tested from a representative subset of vocalizations recorded in the St. Lawrence Estuary, a traditional feeding ground where blue whale migrate in summer and where shipping noise from a major continental seaway represent an additional challenge to automatic detection and identification of low-frequency vocalizations.

2 Chirplet Transform

The chirplet transform (CT) first proposed by Mann and Haykin [7] is an extension of the wavelet transform (WT). It is a time-frequency technique, which decomposes the signal using basis functions, called *Chirplets*. In WT, the basis functions are derived from a single mother wavelet by two operations of dilation and translation [8]. Similarly, the basis of chirplets are derived from a single Gaussian function $g(t) = \pi^{-1/4} \exp(-t^2/2)$ using four operations: scaling, chirping, time and frequency shifting. This leads to a family of chirplets with four adjustable parameters.

$$g_{\tau, f, c, d}(t) = \frac{1}{\sqrt{\sqrt{\pi}d}} e^{-\frac{1}{2}(\frac{t-\tau}{d})^2} e^{j[c(t-\tau)+2\pi f](t-\tau)} \quad (1)$$

where τ , f and c are in \mathbb{R} and d is in \mathbb{R}^+ . The parameters τ , f , c and d represent, respectively, the location in time, the location in frequency, the chirp rate and the duration.

The chirplet transform of a continuous-time signal $x(t)$ is defined by the following inner product:

$$a_{\tau,f,c,d} = \int_{-\infty}^{\infty} x(t)g_{\tau,f,c,d}^*(t)dt \tag{2}$$

where $(*)$ denotes the conjugate complex. The coefficient $a_{\tau,f,c,d}$ represents the signal energy content in a time-frequency region specified by the chirplets [8]. To simplify the notation, the parameter set of the n^{th} chirplet is described by an index set $I_n = \{\tau_n, f_n, c_n, d_n\}$. A given signal $x(t)$ can be approximated as a weighted sum of P chirplets.

$$\hat{x}_P(t) = \sum_{n=1}^P a_{I_n} g_{I_n}(t) \tag{3}$$

where $\hat{x}_P(t)$ is defined as the P^{th} order approximation of the signal.

The optimal estimation of a_{I_n} and I_n corresponding to the decomposition of a signal $x(t)$ into the basis function g_{I_n} can be obtained by the maximum likelihood estimation (MLE) algorithm [8]. In this work, the "DiscreteTFDs" toolbox is used to implement the chirplet transform [9].

3 Application to Simulated Calls

In this section, we apply the CT to simulated A, B and D calls to estimate the parameters under clean and noisy conditions.

3.1 AB Phrases

The A call is simulated using a linear frequency sweeping chirp function $p_{f_A,c_A}(t) = \cos(2\pi(f_A + c_A t)t)$ multiplied by a hanning window $h_{d_A}(t) = \frac{1}{2} + \frac{1}{2} \cos(2\pi t/d_A)$, where f_A is the frequency center, c_A is the chirp rate, and d_A is the duration. In the same manner, the B call is simulated using $p_{f_B,c_B}(t) = \cos(2\pi(f_B + c_B t)t)$ and $h_{d_B}(t) = \frac{1}{2} + \frac{1}{2} \cos(2\pi t/d_B)$ functions. An artificial sound, $x_{AB}(t)$, containing A and B calls, respectively centered on t_{c_A} and t_B , is defined by Eq. 4.

$$\begin{aligned} x_{AB}(t) &= x_A(t) + x_B(t) \\ &= p_{f_A,c_A} \left(t - \tau_A + \frac{d_A}{2} \right) h_{d_A}(t - \tau_A) \\ &\quad + p_{f_B,c_B} \left(t - \tau_B + \frac{d_B}{2} \right) h_{d_B}(t - \tau_B) \end{aligned} \tag{4}$$

The values of the simulated chirp parameters τ , f , c , and d are reported in Tab. 1. In this example, the A and B calls are simulated using two chirps centered on $\tau_A = 25$ s and $\tau_B = 40$ s, respectively.

A noisy version, $y_{AB}(t)$, is obtained by adding white noise, $n(t)$, to $x_{AB}(t)$, with a signal to noise ratio (SNR) of 0 dB.

$$y_{AB}(t) = x_{AB}(t) + n(t) \tag{5}$$

CT is applied to these signals to approximate each simulated call by one chirplet. The chirplets obtained from the clean signal, $\hat{x}_{AB,1}(t) = \hat{x}_{A,1}(t) + \hat{x}_{B,1}(t)$, are an approximation of A and B calls with a single chirplet, respectively. Similarly, chirplets are estimated for the noisy versions, $\hat{y}_{AB,1}(t) = \hat{y}_{A,1}(t) + \hat{y}_{B,1}(t)$. Approximation by one chirplet corresponds to $P = 1$ in Eq. 3.

Fig. 2 shows the simulated A and B calls, $x_{AB}(t)$, and their approximation by two chirplets, and their corresponding spectrograms for the clean and noisy conditions. The estimated CT parameter values of the obtained chirplets are given in Tab. 1.

The chirplet parameters τ , f and c are well estimated, but the parameter d is systematically negatively biased. Values obtained from noisy version are very close to those obtained from the clean signal. CT approximation appears therefore insensitive to noise.

3.2 D Calls

The D call is simulated using a quadratic frequency sweeping chirp $q_{f_D,c_D}(t) = \cos(2\pi(f_D + c_D t^2)t)$ multiplied by a hanning window $h_{d_D}(t) = \frac{1}{2} + \frac{1}{2} \cos(2\pi t/d_D)$, where f_D is the frequency center, c_D is the chirp rate, and d_D is the duration. A signal $x_D(t)$ containing a D call centered on τ_D is generated by Eq. 6.

$$x_D(t) = q_{f_D,c_D} \left(t - \tau_D + \frac{d_D}{2} \right) h_{d_D}(t - \tau_D) \tag{6}$$

The values of the simulated chirp parameters τ_D , f_D , c_D , and d_D are reported in Tab. 2. In this example, a D call is simulated using one quadratic frequency sweeping chirp centered on $\tau_D = 8.5$ s

A noisy version, $y_D(t)$, is obtained by adding white noise $n(t)$ with SNR=0dB.

$$y_D(t) = x_D(t) + n(t) \tag{7}$$

In this section, we used CT to approximate the noisy signal, $y_D(t)$, by one, two and three chirplets. The resulting signals $\hat{y}_{D,1}(t)$, $\hat{y}_{D,2}(t)$ and $\hat{y}_{D,3}(t)$ correspond respectively to $P=1, 2$ and 3 in Eq. 3.

Fig. 3 shows the noisy signal, $y_D(t)$, its approximation by one to three chirplets, $\hat{y}_{D,1}(t)$, $\hat{y}_{D,2}(t)$ and $\hat{y}_{D,3}(t)$, and their corresponding spectrograms. The three-chirplets solution gives the best approximation.

The CT estimation of chirplet parameters (Tab. 2), indicates again that CT approximation is insensitive to noise. For this call, an approximation by a single chirplet gives a relatively accurate estimation of τ and f , but poor estimates of c and d . An approximation by three chirplets is needed to get relatively accurate estimates of all parameters, which can be chosen on this solution as follows: $\tau_D \simeq \tau_{D2} = 8.72$ s, $f_D \simeq f_{D2} = 70.17$ Hz, $c_D \simeq (f_{D3} - f_{D1})/(\tau_3 - \tau_2)^2 = (61.72 - 74.45)/(9.88 - 7.54)^2 = -2.32$ Hz/s and $d_D \simeq d_{D1} + d_{D2} + d_{D3} = 2.28 + 2.23 + 2.01 \simeq 6.52$ s.

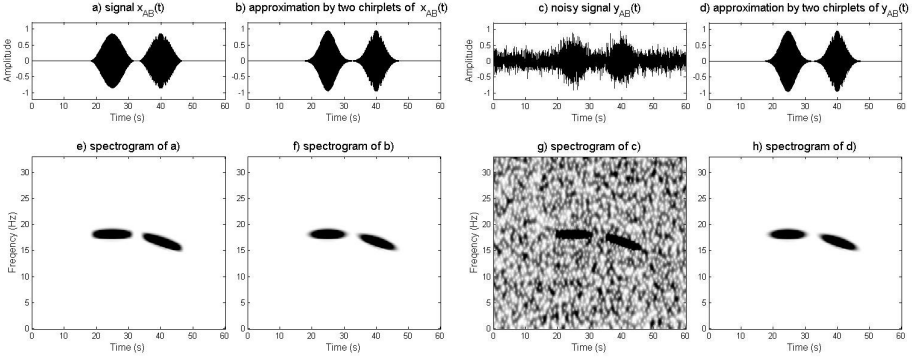


Fig. 2. a) signal $x_{AB}(t)$ representing A and B calls, and b) their approximation by two chirplets, $\hat{x}_{AB,1}(t)$. c) noisy version of A and B calls, $y_{AB}(t)$, with SNR=0dB, and d) its approximation by two chirplets $\hat{y}_{AB,1}(t)$. e-h) represent the corresponding spectrograms.

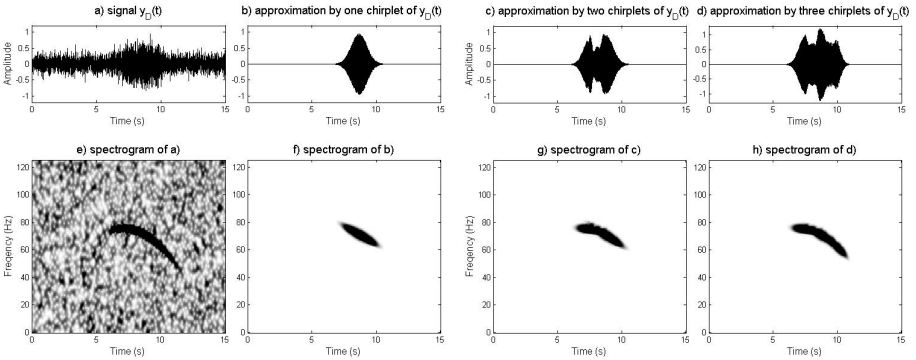


Fig. 3. a) signal representing a noisy simulated D call, $y_D(t)$, using a quadratic frequency sweeping chirp, b) its approximation by one linear frequency sweeping chirplet, $\hat{y}_{D,1}(t)$, c) by two linear frequency sweeping chirplets, $\hat{y}_{D,2}(t)$, and d) by three linear frequency sweeping chirplets, $\hat{y}_{D,3}(t)$, and e-h) represent the corresponding spectrograms

4 Application to Real Calls

In this section, the CT is applied to real blue whale calls to characterize them by a reduced number of parameters.

4.1 Illustrative Example

In this example, the AB phrases and D calls are filtered by 13-23 Hz and 40-80 Hz band-pass filters, respectively. Figs. 4-a,b present the spectrogram of a

Table 1. Chirp parameters estimated by a single chirplet for clean and noisy simulated A and B calls

	Parameter of chirp	Unit	Original value of simulated signal $x_{AB}(t)$	Estimated value of $\hat{x}_{AB,1}(t)$, on clean signal $x_{AB}(t)$	Estimated value of $\hat{y}_{AB,1}(t)$, on noisy signal $y_{AB}(t)$
A call	τ_A	s	25.00	25.00	24.67
	f_A	Hz	18.00	18.00	17.99
	c_A	Hz/s	0.00	0.00	0.00
	d_A	s	14.00	10.46	10.46
B call	τ_B	s	40.00	40.00	40.02
	f_B	Hz	16.50	16.50	16.49
	c_B	Hz/s	-0.21	-0.21	-0.21
	d_B	s	14.00	10.46	10.49

Table 2. Chirp parameters estimated from multiple chirplets on clean and noisy simulated D call

Parameter of simulation	Unit	Original value of simulated signal $x_D(t)$	Estimated value on clean signal $x_D(t)$	Estimated value on noised signal $y_D(t)$
Approximation by one chirplet: $\hat{x}_{D,1}(t)$ and $\hat{y}_{D,1}(t)$				
τ_D	s	8.50	8.63	8.67
f_D	Hz	72.50	70.49	70.25
c_D	Hz/s	-1.60	-5.33	-5.36
d_D	s	7.00	2.73	2.69
Approximation by two chirplets: $\hat{x}_{D,2}(t)$ and $\hat{y}_{D,2}(t)$				
τ_D	s	8.50	7.54 8.79	7.50 8.4
f_D	Hz	72.50	74.46 69.69	74.54 69.97
c_D	Hz/s	-1.60	-1.08 -6.04	-0.97 -6.03
d_D	s	7.00	2.28 2.62	2.11 2.60
Approximation by three chirplets: $\hat{x}_{D,3}(t)$ and $\hat{y}_{D,3}(t)$				
τ_D	s	8.50	7.54 8.72 9.88	7.51 8.72 9.84
f_D	Hz	72.50	74.45 70.17 61.72	74.52 70.15 62.06
c_D	Hz/s	-1.60	-1.19 -5.63 -9.78	-1.09 -5.63 -9.66
d_D	s	7.00	2.28 2.23 2.01	2.12 2.20 1.61

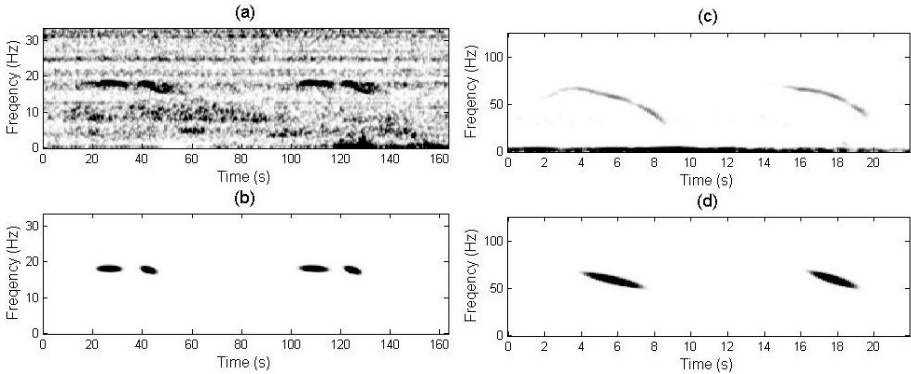


Fig. 4. a) spectrogram of two AB phrases and b) the spectrogram of their approximation with four chirplets. c) Spectrogram of of two D calls and d) the spectrogram of their approximation with two chirplets.

signal containing two AB phrases and their corresponding approximation by four chirplets, one chirplet per call. Figs. 4-c,d show the spectrogram of a signal containing two D calls and their corresponding approximation by two chirplets. The calls are accurately detected despite the presence of noise. Approximation by a single chirplet is sufficient for detecting A and B calls, and is acceptable for D calls.

4.2 Database

The test dataset is constructed from sounds recorded in the St. Lawrence Estuary. The blue whale vocalizations were extracted manually and sorted into three classes: A, B and D. Each class contains 100 calls.

4.3 Feature Extraction

Each individual A, B and D vocalization is approximated by one chirplet as in Fig. 4. The feature vector characterizing an individual vocalization is constructed using the location in frequency f , the chirp rate c and the duration d of the corresponding chirp.

$$v = [f, c, d] \quad (8)$$

where f , c and d are expressed in Hz, Hz/s and s, respectively.

4.4 Vector Quantization

In this project, we used the vector quantization (VQ) to classify the blue whale calls. The VQ is a process of mapping vectors from large space to a finite number of regions in that space. Each region is called a cluster and can be represented by its centroid called *codeword*. For a given vocalization class, the resulting codewords constitute the *codebook* of this class. There are two phases in this classification system: training and recognition. In the training phase, an acoustical model (codebook) is constructed for each vocalization class and the models are stored in a database. In the recognition phase, the unknown vocalization is analyzed and the best matching model is searched from the database. In this experience, we used eight codewords per class.

4.5 Results

To overcome the limited number of calls of the test dataset, the “*k-fold cross-validation*” method is employed to evaluate the performance of the VQ-based classifier. The principle consists in dividing each class into 10 groups, each time we take 9 groups for VQ training and the resting one for the classification test. For each vocalization class, the classification performance is evaluated by the ratio of the number of the correctly classified vocalization (NCCV) to the number of the tested vocalizations (NTV) of this class.

$$\text{Performance (\%)} = \frac{\text{NCCV}}{\text{NTV}} \times 100 \quad (9)$$

Table 3. Performance of VQ-based classifier using three chirplet parameters

Vocalization	Performance (%)
A	93.0
B	97.0
D	98.0
Total	96.0

Results are presented in Table 3. The classification rate was high, with 96% of the calls correctly classified.

5 Conclusion

We presented a new approach based on CT to characterize the blue whale calls. The CT decomposes the signal into gaussian chirplet basic functions with four adjustable parameters, i.e., location in time, location in frequency, chirp rate and duration. Results obtained on simulated and real blue whale calls show that this technique is very promising for efficiently detecting and classifying these calls and robust to noise conditions.

References

1. Sirovic, A., Hildebrand, J.A., Wiggins, S.M.: Blue and fin whale call source levels and propagation range in the southern ocean. *J. Acoust. Soc. Am.* 122, 1208–1215 (2007)
2. Stafford, K.M., Mellinger, D.K., Moore, S.E., Fox, C.G.: Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002. *J. Acoust. Soc. Am.* 122, 3378–3390 (2007)
3. Mellinger, D.K., Clark, C.W.: Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic. *J. Acoust. Soc. Am.* 114(2), 1108–1119 (2003)
4. Berchok, C.L., Bradley, D.L., Gabrielson, T.B.: St. Lawrence blue whale vocalizations revisited: Characterization of calls detected from 1998 to 2001. *J. Acoust. Soc. Am.* 120, 2340–2354 (2006)
5. Mellinger, D.K., Stafford, K.M., Moore, S.E., Dziak, R.P., Matsumoto, H.: An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography* 20(4), 36–45 (2007)
6. Mellinger, D.K., Clark, C.W.: Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acoust. Soc. Am.* 107(2), 3518–3529 (2000)
7. Mann, S., Haykin, S.: Adaptive Chirplet Transform: an adaptive generalization of the wavelet transform. *Optical Engineering* 31(6), 1243–1256
8. Cui, J., Wong, W.: The adaptive chirplet transform and visual evoked potentials. *Biomedical Engineering, IEEE Transactions on* 53(7), 1378–1384 (2006)
9. J.C. O'Neill, DiscreteTFDs: Time-Frequency Analysis Software, <http://tfd.sourceforge.net/>