A standardized method of classifying pulsed sounds and its application to pulse rate measurement of blue whale southeast Pacific song units

Julie Patris,[1,a)] Franck Malige,[1] Hervé Glotin,[1] Mark Asch,[2] and Susannah J. Buchan[3,b)]

[1]Université de Toulon, l'Université d'Aix-Marseille, Centre National de la Recherche Scientifique, Laboratoire d'Informatique et Systèmes, DYNamique de l'Information Team, Campus de La garde, 83130, France
[2]Université de Picardie Jules Verne, Centre National de la Recherche Scientifique, Laboratoire Amiénois de Mathématique Fondamentale et Appliquée, Chem Thil, 80025 Amiens Cedex 1, France
[3]Center for Oceanographic Research in the Eastern South Pacific Sur-Austral, Edificio Departamento de Oceanografía Piso 2 Barrio Universitario s/n, Universidad de Concepción, Concepción, Chile

Pulsed sounds are an interesting example of complex biological sounds, frequent in cetaceans' vocalizations. A two-group classification of these sounds is proposed: tonal or non-tonal. Through two simple mathematical models, it is shown that this classification can lead to better techniques for measuring the pulse rate. This classification is thus useful for improving measurement accuracy, but can also help in formulating hypotheses regarding mechanisms of sound production. This method of classification is applied to south Pacific blue whale vocalizations and it is found that the pulse rate corresponds to the fundamental frequency (not expressed in the spectrum) of the song. Thus, the hypothesis that the sound is produced by only one organ and then filtered by the body of the giant is reinforced. © 2019 Acoustical Society of America. https://doi.org/10.1121/1.5126710

I. INTRODUCTION

A. Pulsed sounds in blue whale song types

Among the numerous varieties of animal sounds, pulsed sounds are particularly complex and interesting and are common in marine mammals (Au et al., 2000; Richardson et al., 1995). A pulsed sound is the repetition of similar "pulses" or short signals, with a constant pulse rate, and aurally these sounds are often perceived by humans as amplitude modulated sounds. Among cetaceans, such a repetition of similar "pulses" has been studied in odontocete vocalizations (Murray et al., 1998; Miralles et al., 2012) and is also described in mysticete songs (Cummings and Thompson, 1971). By way of an example, two biological pulsed sounds from marine mammals—blue whale (*Balaenoptera musculus*) and bottlenose dolphin (*Tursiops truncatus*)—are represented in the time domain in Fig. 1.

As with most baleen whales, blue whales produce high energy, low frequency, and long duration vocalizations (Cummings and Thompson, 1971) that are highly structured in time, with endless repetition of remarkably self similar phrases. Since only males have been reported to produce these sounds (Oleson et al., 2007), they are thought to play a role in reproduction, as happens in humpback whales (*Megaptera novaeangliae*) (Glockner, 1983). Several very distinct song types have been registered for blue whales worldwide, each characteristic of a population, and most of these song types include units that are aurally characterized as pulsed (McDonald et al., 2006). Because of the very interesting phenomenon of frequency decrease in blue whales' songs (McDonald et al., 2009), it is particularly critical to have good descriptors of the song, and the pulse rate might be a useful metric. However, though the pulse rate is generally part of the description of the song (Watkins, 1968, Cummings and Thompson, 1971; Miller et al., 2014), no recent efforts have been made towards studying its properties and the best methods to measure it.

In this paper we aim to show that it can be useful to classify pulsed sounds in two different classes: a tonal (or periodic) signal, or a non-tonal (aperiodic) one: first, because the technique for measuring the pulse rate is different whether the signal is tonal or not; and second, because it can help in understanding the sound production mechanisms.

B. Sound production in cetaceans

Sound production in large marine mammals is a difficult subject since live animals cannot be examined. Though physical and anatomical analyses have been proposed (Aroyan et al., 2000), sound production in mysticetes is still poorly known compared to odontocetes (Au et al., 2000) and is an active area of research (Mercado III et al., 2010; Reidenberg, 2017). Some studies try to reproduce sound production using theoretical models or sound production machines (Adam et al., 2013). According to these studies and anatomical analyses, sound could be produced by vibrating U-shaped vocal folds and a complex system of resonators (lungs, laryngeal sac, trachea, and other tissues) that modify the sound as a passive filter.
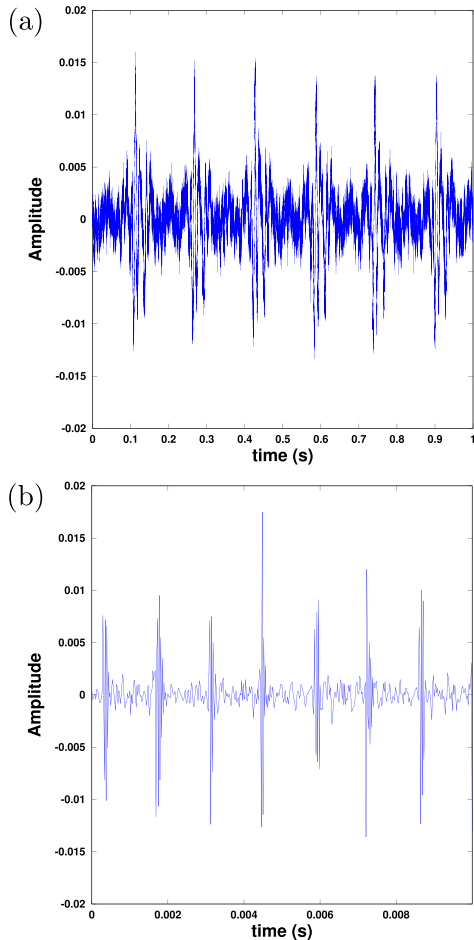
FIG. 1. (Color online) Waveform of two biological pulsed sounds, both recorded off Chañaral de Aceituno Island, Chile, in 2017 with an autonomous recorder at $f_s = 48$ kHz (see Sec. III B for details on data acquisition). The amplitude on the y axis is in relative pressure. (a) Extract of unit B of the southeast Pacific song type 2 of a blue whale (*Balaenoptera musculus*). (b) Extract of a bottlenose dolphin (*Tursiops truncatus*) buzzed sound. The signals have been identified by experts, given the ecological context of the location, and through comparison with other recordings made locally.

Regarding pulsed sounds, they can be produced through the action of two organs as in killer whales (*Orcinus orca*) (Brown, 2008). In this case the sound produced is called "two-voiced sound": one high frequency sound is modulated in amplitude by the action of another organ at a much lower frequency.

Another way to produce pulsed sound is to use one organ which regularly produces similar pulses. Reidenberg (2017) proposed such an organ for the production of pulses in mysticetes' sounds: "a pair of broad tissue flaps supported by the corniculate cartilages." Dziak *et al.* (2017), in a recent study of north Pacific blue whales' songs, also shows the growing interest in finding clues in signal processing and signal modeling for sound production.

C. Classification of pulsed sound for analysis

The goal of this paper is to propose a simple classification of pulsed sound, along with two mathematical models (Sec. II). We show the interest of such methods for obtaining a better description of the sound, to achieve more accurate
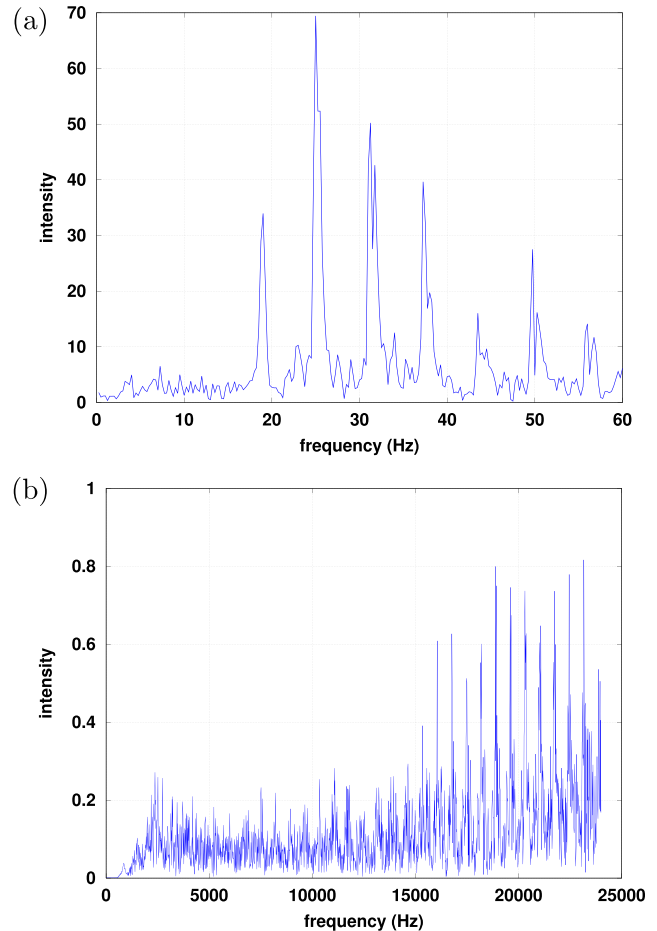
FIG. 2. (Color online) Spectra (by means of a FFT) of the two biological pulsed sounds presented in Fig. 1. (a) Unit B of the southeast Pacific song type 2 of a blue whale (*Balaenoptera musculus*), signal length is 4 s. (b) Buzzed sound of a bottlenose dolphin (*Tursiops truncatus*), signal length is 0.07 s.

measurements of marine mammals' vocalizations, but also as a clue to sound production mechanisms (Sec. II). While most of the efforts of describing pulsed sounds have been dedicated to odontocetes (Murray *et al.*, 1998; Rankin *et al.*, 2007; Simard *et al.*, 2008; Brown, 2008), we propose to analyze pulsed sounds emitted by blue whales in Chilean coastal seas (Buchan *et al.*, 2014) as an example of the application of our method (Sec. III). In this case, we show that our analysis gives a better precision in the songs' description and tends to confirm the source-filter hypothesis for sound production.

II. A METHOD FOR THE ANALYSIS OF CETACEANS' PULSED SOUNDS

A. Tonal sound versus non-tonal sound

The Fourier transform of a pulsed sound shows peaks (or lines) of frequencies, with a constant separation between frequencies (see Fig. 2). As can be verified in our models below, and is also shown in Watkins' original paper (Watkins, 1968), this band interval corresponds, in the time domain, to the repetition rate of pulses, or pulse rate, called f_{pulse} in our study. These bands are often called side-bands in the literature. In order to better understand and characterize

TABLE I. First peak frequencies and average pulse rate, in Hz, of the two examples shown in Fig. 2 as measured from the spectra. Uncertainties (measured by the standard deviation, σ) are 0.25 Hz for blue whale and 15 Hz for bottlenose dolphin frequency measurements. Standard deviations for Δf have been computed by measuring all peaks and not only the first four.

	f_1	f_2	f_3	f_4	Δf	$\sigma(\Delta f)$
Whale	19.0	25.2	31.3	37.6	6.26	0.25
Dolphin	18900	19610	20330	21040	712	6

the pulsed sound, we will examine the relation between this side-band separation (or pulse rate) and the abscissa of the peaks.

Let us call $\{f_i\}$ the abscissa of the peaks in the spectrum and $\Delta f = f_{\text{pulse}}$ the constant band interval and visualize them in our examples. Table I gives the peak frequencies $\{f_i\}$ and the averaged band interval or pulse rate $\Delta f = f_{\text{pulse}}$ for the two examples of pulsed sound presented in Fig. 2. Frequency uncertainty due to signal length limitation is 0.25 Hz for the blue whale's signal and 15 Hz for the dolphin's.

We propose the following criterion for a classification of pulsed sounds in two groups: if there exists $(k_i) \in \mathbb{N}$ such that for all i,

$$f_i = k_i \Delta f \tag{1}$$

then the sound will be called tonal (although the fundamental frequency may not be visible or expressed in the spectrum). That is, the peak frequencies $\{f_i\}$ are all integer multiples of the pulse rate Δf. In this case the signal is periodic of period $T_{\text{pulse}} = 1/f_{\text{pulse}}$, and f_{pulse} is the fundamental frequency of the sound. Else, the pulsed sound will be called non-tonal. If we examine again our two examples of the blue whale song and the dolphin buzz, we see that the former can be classified as a tonal signal, whereas the latter cannot (see Table II).

Based on this classification, Sec. II B will examine how the pulse rate can be measured in each case, with the help of two mathematical models.

B. Mathematical models for the measurement of pulsed sound parameters

In this section we present two mathematical models of pulsed sounds. The first one, model A, can only apply to a signal where the peak frequencies $\{f_i\}$ are integer multiples of the band interval or pulse rate Δf (tonal signals). The second one, model B, is more general and can be applied to either of the two possibilities described in Sec. II A. In both

TABLE II. Ratio between frequencies f_i of Table I and pulse rate of the two examples shown in Fig. 2. A ratio close to an integer is compatible with a tonal sound. A non-integer ratio excludes the possibility of having a tonal sound.

	$\frac{f_1}{\Delta f}$	$\frac{f_2}{\Delta f}$	$\frac{f_3}{\Delta f}$	$\frac{f_4}{\Delta f}$
Whale	3.04 ± 0.1	4.03 ± 0.1	5.00 ± 0.1	6.01 ± 0.1
Dolphin	26.55 ± 0.24	27.54 ± 0.24	28.55 ± 0.24	29.55 ± 0.24

models, we consider the pulsed sound as infinite in time, which means we are not addressing the effects of the global duration of the sound. Eventually, if needed, a window w corresponding to the global duration of the signal will be used in the computation of theoretical formulas. For each model we present and compute the Fourier transform and autocorrelation function of the signal. These two operators are often used in signal processing to analyze the signals and to measure parameters such as peak frequency and pulse rate.

1. Model A

Let us first consider a pulsed sound as the repetition of distinct and similar pulses, separated by a duration T_{pulse}. This is the point of view developed by Dziak *et al.* (2017) to model unit *B* of the northeast Pacific blue whale song type. Then the easiest way to mathematically represent such a function of time is the convolution of a specific wavelet p (the pulse) by a Dirac comb $\text{III}_{T_{\text{pulse}}}$, characterized by the time T_{pulse} between each impulse. We note $f_{\text{pulse}} = 1/T_{\text{pulse}}$.

Thus, the signal shown in Fig. 3 can be written as

$$s_A(t) = \sum_{n \in \mathbb{Z}} p(t - nT_{\text{pulse}}) = \left[p * \text{III}_{T_{\text{pulse}}} \right](t), \tag{2}$$

where $*$ is the convolution symbol, $[g * h](t) = \int_{-\infty}^{+\infty} g(u)h(t-u)du$, $\text{III}_{T_{\text{pulse}}}(t) = \sum_{n \in \mathbb{Z}} \delta_{nT_{\text{pulse}}}(t)$ is the Dirac comb distribution of period T_{pulse}, and $\delta_{t_0}(t)$ is the Dirac distribution centered at t_0. A good reference on these techniques is Appel (2008). We note that s_A is then a periodic function, of period T_{pulse}. There is no phase difference between the pulses (see Fig. 3). A more complete model could include an additive noise term, $\nu(t)$, usually assumed to be of zero mean and known (estimated) variance.

The Fourier transform of the function $s_A(t)$ is defined by $S_A(f) = \int_{-\infty}^{+\infty} s_A(t)e^{-2i\pi ft}dt$ and gives

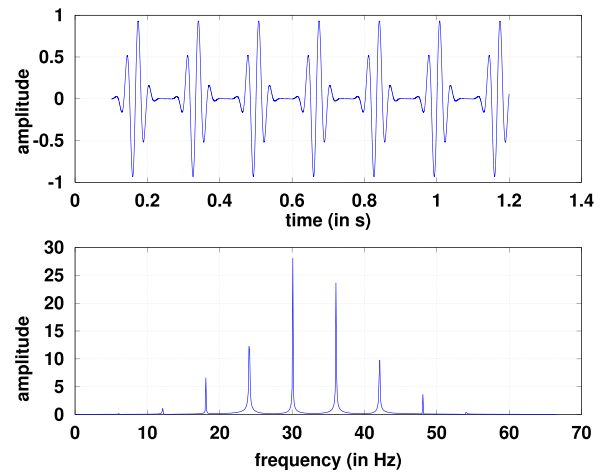$$S_A(f) = P(f) \times \text{III}_{f_{\text{pulse}}}(f), \tag{3}$$

FIG. 3. (Color online) Model A in waveform (top) and its FFT (bottom). In model A, we choose the pulse as the product of a sine function of frequency $f_0 = 31.7$ Hz with a Gaussian of standard deviation ($\sigma = 0.02$ s). We also choose $f_{\text{pulse}} = 6$ Hz. In the FFT we can see that the peaks are multiples of $f_{\text{pulse}} = 6$ Hz.

where $P(f)$ is the Fourier transform of the wavelet. In Fig. 3, $p(t)$ is a Gaussian multiplied by a sine of frequency $f_0 = 31.7$ Hz and its Fourier transform $P(f)$ is therefore a Gaussian centered on f_0. We observe that the Fourier transform of the signal is the spectrum of the pulse P multiplied by a Dirac comb. Thus, the spectrum of s_A is a set of frequency bands at integer multiples of f_{pulse} (Fig. 3, bottom). The frequency band with higher energy (30 Hz) does not correspond to $f_0 = 31.7$ Hz (even if the wavelet energy was maximal at this frequency) because of the multiplication by the Dirac comb.

It is important to underline that in practice, the signals analyzed are finite of duration T_{signal}. In this case, we can write $s_{A,\text{finite}}(t) = s_A(t) \times w(t)$ where w is a window of duration T_{signal}. A classical window is the rectangular window, built on an indicator function $w(t) = \mathbb{1}_{[-T_{\text{signal}}/2;T_{\text{signal}}/2]}(t)$, but any kind of window can be used, as in Dziak *et al.* (2017). In this case, the Fourier transform is

$$S_{A,\text{finite}}(f) = \left[(P \times \text{III}_{f_{\text{pulse}}}) * W\right](f), \quad (4)$$

where W is the Fourier transform of w. In the case of a rectangular window, the Fourier transform is $W(f) = T_{\text{signal}} \times \text{sinc}(\pi T_{\text{signal}} f)$, a cardinal sine giving a width to the peaks in Fig. 3, linked to the value of T_{signal}.

The autocorrelation function of a signal s is $C_s(\tau) = \lim_{T \to +\infty} (1/T) \int_{-T/2}^{T/2} s(t)s^*(t+\tau)dt$, where s^* is the complex conjugate of s. In the case of a finite signal $s_{A,\text{finite}}$ and a rectangular window $w(t) = \mathbb{1}_{[-T_{\text{signal}}/2;T_{\text{signal}}/2]}(t)$, the autocorrelation function is

$$C_{s_{A,\text{finite}}}(\tau) \simeq \Lambda\left(\frac{\tau}{T_{\text{signal}}}\right) \times \left(\sum_{n \in \mathbb{Z}} |P(nf_{\text{pulse}})|^2 e^{2i\pi nf_{\text{pulse}}\tau}\right), \quad (5)$$

where $\Lambda(t)$ is the triangular function ($\Lambda(t) = 1 + t$ on [–1; 0], $\Lambda(t) = 1 - t$ on [0,1] and zero outside of [–1; 1]). The proof is very similar to the proof given in the Appendix for model B and will not be detailed here. An important remark is that the first maximum of the modulus of the autocorrelation function (other than $\tau = 0$) is obtained for $\tau = T_{\text{pulse}}$, the period of the signal. Thus, for this model of pulsed sound, the autocorrelation, or the summed autocorrelation (Wise *et al.*, 1976), is a good, unbiased tool to measure the pulse rate.

2. Model B

We will now examine the case when the pulsed sound can be described as a tonal sound modulated in amplitude by a periodic function. This kind of pulsed sound has been described by Watkins (1968) or Brown (2008). A straightforward way to represent this signal is to multiply a tonal function g_{T_0} (characterized by a fundamental frequency f_0) by a signal that could be an envelope e convolved with a Dirac comb (of period T_{pulse}). We assume that $f_{\text{pulse}} \ll f_0$ so that the tonal function g_{T_0} is modulated in amplitude by a function with a much smaller frequency.
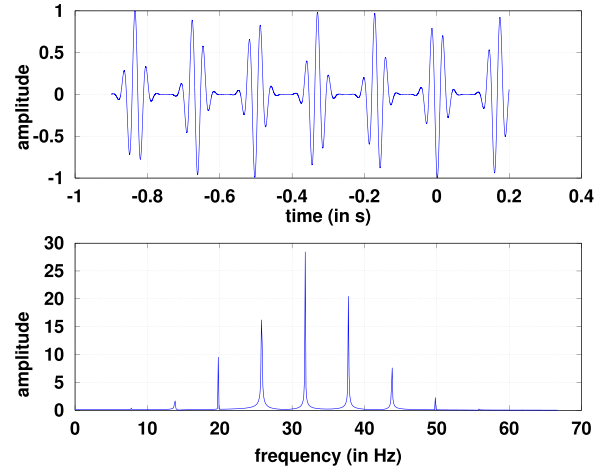
FIG. 4. (Color online) Model B in waveform (top) and its FFT (bottom). In this model, we choose the tonal function g_{T_0} as a pure sine function of period T_0 and the envelope e as a Gaussian with standard deviation $\sigma = 0.02$ s. As in Fig. 3, we choose $f_0 = 31.7$ Hz and $f_{\text{pulse}} = 6$ Hz. In the FFT we can see that the frequency peaks are not centered at multiples of f_{pulse}.

We write

$$s_B(t) = g_{T_0}(t) \times \sum_{n \in \mathbb{Z}} e(t - nT_{\text{pulse}})$$

$$= g_{T_0}(t) \times \left[e * \text{III}_{T_{\text{pulse}}}\right](t). \quad (6)$$

In this case, the signal is not a periodic function. If we examine each of the "pulses," they do not have the same phase (see Fig. 4). This is due to the multiplication of two tonal functions with different periods.

Let us compute the Fourier transform of such a signal. We obtain

$$S_B(f) = \sum_{n \in \mathbb{Z}} E(nf_{\text{pulse}})G_{f_0}(f - nf_{\text{pulse}}), \quad (7)$$

where E and G_{f_0} are the Fourier transforms of e and g_{T_0}, respectively. The proof is given in the Appendix. In this formula, as g_{T_0} is a tonal sound, only multiples of f_0 are found in its spectrum and thus $S_B(f)$ is different from zero only if $f - nf_{\text{pulse}}$ is a multiple of f_0, that is only if $f = nf_{pulse} + mf_0$, which is usually not a multiple of f_{pulse}. In this case, sidebands appear in the spectrum at distances nf_{pulse} of mf_0. In addition, if the tonal signal $g_{T_0}(t)$ has its energy concentrated at the frequency f_0, and if the Fourier transform of the envelope $E(f)$ is sufficiently regular, we see that the pulsed signal s_B has a maximum of energy also at f_0 (see Fig. 4).

As g_{T_0} is a tonal signal with fundamental f_0, we can write its Fourier series as $g_{T_0}(t) = \sum_{n \in \mathbb{Z}} a_n e^{2i\pi nf_0 t}$. Then the autocorrelation function of the finite signal associated to model B is

$$C_{s_{B,\text{finite}}}(\tau) \simeq \Lambda(\tau/T_{\text{signal}}) \left(\sum_{n \in \mathbb{Z}} |a_n|^2 e^{2i\pi nf_0 \tau}\right)$$

$$\times \left(\sum_{m \in \mathbb{Z}} |E(mf_{\text{pulse}})|^2 e^{2i\pi mf_{\text{pulse}}\tau}\right). \quad (8)$$

The proof is given in the Appendix. Contrary to the model A, the non-zero maximum of this function is not obtained for $\tau = T_{\text{pulse}}$ (see remark in the Appendix). Thus, the

maximum of the autocorrelation function is a biased estimator of the pulse rate in this case.

C. Conclusions on the method

1. Consequences on measurement techniques

As can be seen in Sec. II B, if the pulsed sound is tonal, the autocorrelation function is a good tool to measure the pulse rate f_{pulse}. This can be of importance because the autocorrelation function method can be much more precise than other methods, as is shown in an example in Sec. III E.

Thus, our recommendation for a precise measurement of pulsed sound characteristics would be to:

(1) compute the fast Fourier transform (FFT) of the whole pulsed signal (the FFT resolution in frequency is $1/T_{\text{signal}}$, so it is important to have as long a signal as possible);
(2) measure the frequencies, $\{f_i\}$, corresponding to peaks in the FFT;
(3) compute their interval $\Delta f = f_{\text{pulse}}$ or find an approximation of $\Delta f = f_{\text{pulse}}$ by getting the envelope of the signal (the envelope can be obtained by squaring the signal and low-pass filtering it) and then compute the maximum of the autocorrelation function of the envelope;
(4) check whether the signal is tonal or not by examining the quotients $f_i/\Delta f$;
(5) if the signal is tonal, get a better approximation of $\Delta f = f_{\text{pulse}}$ by finding the first maximum of the modulus of the summed autocorrelation function.

2. Consequences on sound production

If the sound is tonal (as in model A for example), then it is compatible with the action of only one source of sound production, altered by a passive filter (source-filter theory) (Fant, 1960). This is the same case as formants in human voice production (Flanagan, 1965; Howard and Angus, 2006), but the sound will appear "pulsed" or "amplitude modulated" when the first harmonics are filtered out. This has been shown to be the case in some musical instruments (e.g., timpani or trombone) or some birds [e.g., *Parus atricapillus* (Nowicki and Capranica, 1986)]. Tonal pulsed sounds are also compatible with the regular repetition of very short separated pulses produced by one organ as proposed by Dziak *et al.* (2017) or by Reidenberg (2017).

If the sound is not tonal (as in model B for instance), then it is a combination of sounds with two different frequencies (f_0 and f_{pulse}) that are not linked. In this case, it is not compatible with the source-filter theory. One possibility is that two independent organs are used to produce the sound: one produces a signal, and the other acts as an amplitude modulation of the first signal. As shown in Sec. II A, the dolphin's buzz seems not to be a tonal signal, and our analysis is thus consistent with the mechanism proposed by Au *et al.* (2000) (p. 114) to explain rapid trains of clicks involving the concomitant action of two generators.

III. APPLICATION TO BLUE WHALE SONG UNITS

A. The southeast Pacific blue whale song type

In this paper we are interested in a southeast Pacific blue whale song called SEP2, first recorded in 1996 (Stafford *et al.*, 1999) and first described in detail in 2014 (Buchan *et al.*, 2014). A representation of the repeated phrase is given in Fig. 5. This phrase, composed of several units, is usually repeated every two minutes in a sequence lasting from some minutes to a few hours.

These units have been described as pulsed (Buchan *et al.*, 2014). Their spectra show a set of peaks that are regularly spaced but where the fundamental frequency does not appear. This aspect is not due to a propagation effect, since all recordings of this song type show the same aspect
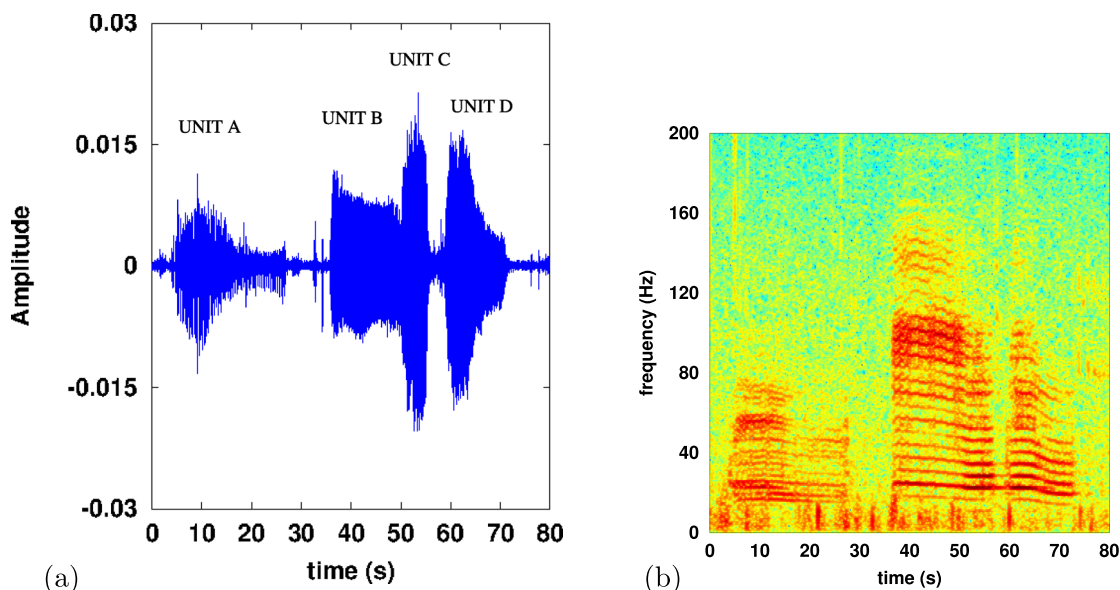

FIG. 5. (Color online) Phrase of the southeast Pacific blue whale song, SEP2, recorded off Isla Chañaral, Chile, February 2, 2017, sample frequency f_s = 48 kHz. (a) Waveform of relative intensity. (b) Time-frequency representation (FFT 2^{12} points, overlap of 90%, Hanning window). Low frequency bars are background noise.

J. Acoust. Soc. Am. **146** (4), October 2019

Patris *et al.* 2149

independently of the place and technology of the recording device (Stafford *et al.*, 1999; Buchan *et al.*, 2014). Alternatively, pulses are also visible if we zoom in on the waveform, as in Fig. 1, top. However, the amplitude modulation visible on the waveform is not rectangular as in separated pulses, but rather like a sinusoidal modulation.

B. Data collection

Data were collected close to the Isla Chañaral marine reserve in northern Chile, between the Isla Chañaral and the mainland, at $29°00'44''$ south and $71°31'26''$ west during the austral summer of 2016/2017, between 16 January 2017 and 27 February 2017. The hydrophone and recording package BOMBYX II was deployed at 15–20 m below the surface on a mooring where water column depth was 70 m. Data were collected during three periods of two weeks in January and February (Patris *et al.*, 2017). The recording device BOMBYX II was mounted by the University of Toulon and includes a Cetacean Research C57 hydrophone and a commercial SONY PCM-M10 recording device. The hydrophone has very high sensitivity, a flat response down to 20 Hz and is omni directional at low frequencies (less than 10 kHz). The hydrophone is alimented by 9 V and high-pass filtered (C = 47 μF, frequency cut 0.15 Hz) to remove the continuous component. The SONY PCM-M10 recorder (gain 6, Rin = 22 kOhm) is equipped with a 256 GB memory card, and set up in a specialized tube made by Osean able to resist high pressure (Patris *et al.*, 2017). Recording was done at a sample rate of 48 kHz so as to record a vast diversity of cetaceans, ranging from large whales to dolphins (namely bottlenose dolphins, *Tursiops truncatus*), and at 16-bits, allowing for high sensitivity without saturating the memory.

C. Analysis

A systematic analysis showed that blue whales' songs were present almost all days of recording (Balcazar and Alosilla, 2018). Long series of up to 70 phrases of high signal to noise ratio (SNR) were recorded, especially on Feb, 2nd 2017.

We selected 100 high SNR phrases from this data set. These phrases were selected visually on waveforms or spectrograms when all four units are clearly seen (Fig. 5). The phrases were selected on six different days of our recording to reduce the probability to have only one individual producing the songs. On these phrases, we decided to apply our criterion to

characterize the nature of these blue whales' "pulsed" sounds. To this end, we measured the peak frequency set $\{f_i\}$ and pulse rate $\Delta f = f_\text{pulse}$. For the selected high SNR signals, we analyzed the four units A, B, C, and D of the signals (see Fig. 5) that have different frequency characteristics but are all pulsed. These units are described in detail in Buchan *et al.* (2014).

1. Peak frequency

For all selected units, we performed an FFT on the first 4 s of the unit by a routine in OCTAVE (Eaton *et al.*, 2009). We measured one of the peak frequencies, which is the one that on average shows the highest SNR. This f_i is measured as the frequency corresponding to the maximum value (in modulus) of the FFT between 23 and 25 Hz for unit A and between 22 and 26 Hz for units B, C, and D. As we performed an FFT on $T_\text{signal} = 4$ s of the signal, there is a quantification of the measure of the frequency which is equal to $1/T_\text{signal} = 0.25$ Hz and thus the uncertainty on this measure is on the order of 1%. Due to the fact that the precision in frequency is inverse to the duration of the signal, it is important to use as long a signal as possible. Nevertheless, for consistency, we chose to extract 4 s of each song unit. Indeed, unit A, B, C, and D all last longer than, or exactly 4 s in each recording.

2. Pulse frequency

The estimation of f_pulse by a difference of two frequencies $\{f_i\}$ obtained by the FFT would lead to a poor precision, on the order of 8%. Thus, to measure the pulse rate f_pulse of the signal with a better precision, we first performed an envelope detection. To this effect we squared the signal and then low pass filtered it using a fifth order Butterworth filter with frequency cut-off at 10 Hz. Other methods of reconstructing the envelope of the signal can be used (Glotin, 2001) giving similar results. Then a summed autocorrelation (Wise *et al.*, 1976) on the first 4 s of the signal was performed to measure the pulse rate. The relative uncertainty on this measure is around 1.5% (see Sec. III E).

D. Results

The results of the measures of the ratio between f_i and f_pulse for the four units of the SEP2 phrase are shown in Fig. 6.

As we can see, the dispersion of the ratio value f_i/f_pulse around a fixed integer number is small, especially for units C
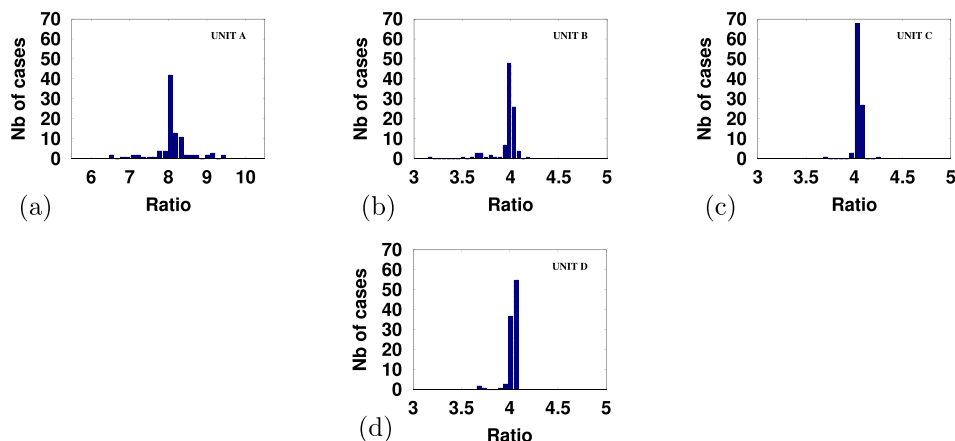

(a)

(b)

(c)

(d)

FIG. 6. (Color online) For 100 high SNR SEP2 phrases in 2017, histograms of the ratio between the peak frequency, measured by an FFT, and the pulse rate, measured by envelope detection and summed autocorrelation for units A, B, C, and D [(a), (b), (c) and (d), respectively].

2150   J. Acoust. Soc. Am. **146** (4), October 2019

Patris *et al.*

TABLE III. Mean fundamental frequencies f_0 (shown to be the same as the pulse frequency) with standard deviation of the four units of the SEP2 song for the 100 phrases recorded in 2017.

Unit	A	B	C	D
f_{pulse} (Hz)	2.98 ± 0.19	6.52 ± 0.17	5.88 ± 0.08	5.89 ± 0.11

and D, which usually have a better signal to noise ratio. This dispersion can be explained by errors in measurements (see Sec. III C), presence of additional low-frequency noise (see Fig. 5) or variability in the frequency (especially for unit B). For unit A, the ratio is near 8, and for the other units near 4. Thus, our measures are compatible with the hypothesis of a tonal signal for the four units of the SEP2 song phrases. The values of the very low fundamental frequencies f_0 (which coincides with the pulse rate f_{pulse}) are given in Table III for 2017. This fundamental frequency is very stable between two phrases.

E. Discussion

1. Measurements of pulse rate

As seen in Sec. II C, in the case of a tonal pulsed sound, the measurement of f_{pulse} can be done without bias by at least three different methods: FFT of the signal and measure of the gap between two frequency peaks as in Sec. II A, summed autocorrelation of the envelope of the signal (see Sec. III C), summed autocorrelation of the signal (Wise et al., 1976). In Fig. 7, we present three histograms of the values of f_{pulse} (for unit C) measured by these three methods on our set of 100 signals.

In the case of the measure of the difference between two peaks of the FFT, the result has a quantification value of 0.25 Hz. This value is clearly seen in Fig. 7 and this method is inefficient for measuring f_{pulse} in this configuration (short duration of the signal compared to the pulse period T_{pulse}). The mean value and standard deviation of the measure is in this case $f_{pulse} = 5.9 \pm 0.2$ Hz. However, the statistical distribution of the values is far from being a normal distribution [see Fig. 7(a)], so the standard deviation is clearly not a tool that is adapted to this result.

In the case of the autocorrelation of the envelope of the signal, we obtain $f_{pulse} = 5.88 \pm 0.08$ Hz [see Fig. 7(b)]. In the case of the summed autocorrelation, we obtain $f_{pulse} = 5.88 \pm 0.02$ Hz [see Fig. 7(c)]. The best precision is thus obtained by summed autocorrelation of the signal, however, this measurement technique can only be used for a tonal signal. Thus, we see the importance of first assessing the tonal nature of the signal before measuring the pulse rate.

2. Sound production

Since we have shown that SEP2 units are tonal, they are compatible with a sound produced by only one organ. The whale song is either produced by a source-filter mechanism [that could be U-folds and the body of the animal as described in Reidenberg (2017)], or by the regular repetition of pulse production by an organ (Dziak et al., 2017). In each case the fundamental frequency of vibrations is given by the pulse rate and is around 3 or 6 Hz for SEP2 units.

This could also be the case for another type of blue whale song, units B and C of the SW Pacific song type recorded off
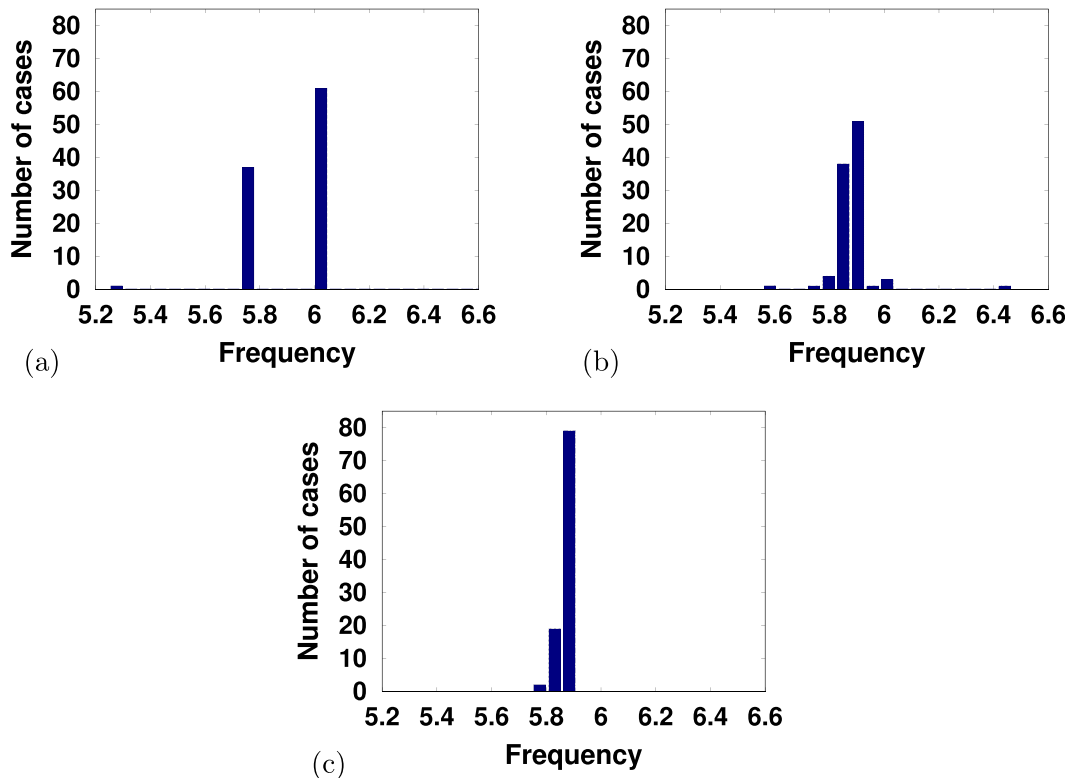

(a)

(b)

(c)

FIG. 7. (Color online) Histograms of the estimation of pulse rate f_{pulse} for unit C of 100 high SNR blue whale phrases. Three different methods of estimation were employed. (a) Estimation by difference of frequency peaks f_i in the FFT. (b) Estimation by summed autocorrelation of the envelope of the signal. (c) Estimation by summed autocorrelation of the signal.

New Zealand. Indeed, in Miller *et al.* (2014), a joint yearly decrease at almost the same relative rate of peak and pulse frequencies tends to show that these frequencies are linked, probably by an integer ratio, f_{peak} being around ten times f_{pulse}.

However, pulsed units of blue whale song types worldwide seem to not always be tonal sounds as are SEP2 units. For example, the first unit of the pygmy blue whale song type from south and west Australia (SE Indian song type) has side bands indicating its pulsed nature (Gavrilov *et al.*, 2011; Stafford *et al.*, 2011). Spectrograms presented in these studies show that during a song occurrence the peak frequency increases (upsweep) while the pulse rate decreases (the gap between frequency side-bands narrows) (Watkins, 1968). Thus, these two frequencies are not linked by an integer ratio, and their production is probably not explained by a source-filter mechanism.

IV. CONCLUSION

In this paper we presented mathematical models to better characterize and thus understand a pulsed sound. Though we compared them to only one example, the SEP2 blue whale phrase, they could be compared to numerous other biological pulsed sounds. The four units of the SEP2 song type are found to be tonal in spite of their pulsed nature. Thus, the fundamental frequency (or pulse rate) is the best tool to characterize it. Some studies, like the study of frequency decline in baleen whales (McDonald *et al.*, 2009), compare several song types worldwide. It would be useful to have a single, common criterion to characterize a sound, and the fundamental frequency may be a good option.

APPENDIX: COMPUTATION OF THEORETICAL FORMULAS AND PROOFS

In this appendix we present proofs of the results stated in Sec. II B.

1. Fourier transform of the model B

Statement: If $s_B(t) = g_{T_0}(t) \times [e * \text{III}_{T_{pulse}}](t)$, then $S_B(f) = \sum_{n\in\mathbb{Z}} E(nf_{pulse}) G_{f_0}(f - nf_{pulse})$ is its Fourier transform.

Proof:

$$S_B(f) = \left[G_{f_0} * (E \times \text{III}_{f_{pulse}})\right](f)$$
$$= \left[\sum_{n\in\mathbb{Z}} E(nf_{pulse})\delta(f - nf_{pulse}) * G_{f_0}\right](f)$$
$$= \sum_{n\in\mathbb{Z}} E(nf_{pulse}) G_{f_0}(f - nf_{pulse}).$$

2. Autocorrelation function of the model B

The signal $s_{B,finite} = g_{T_0}(t) \times [e * \text{III}_{T_{pulse}}](t) \times w(t)$ is of the form

$$s_{B,finite} = \sum_{n\in\mathbb{Z}} a_n e^{2i\pi n f_0 t} \times \left(\sum_{n\in\mathbb{Z}} e(t - nT_{pulse})\right) \times w(t)$$

considering that g_{T_0} is a tonal sound with fundamental equal to f_0 and thus can be expressed as $\sum_{n\in\mathbb{Z}} a_n e^{2i\pi n f_0 t}$.

Statement: Let a finite pulsed sound

$$s_{B,finite}(t) = \left(\sum_{n\in\mathbb{Z}} e(t - nT_{pulse})\right) \times \sum_{n\in\mathbb{Z}} a_n e^{2i\pi n f_0 t} \times w(t),$$

where $w(t) = \mathbb{1}_{[-T_{signal}/2; T_{signal}/2]}(t)$ which satisfies the two hypotheses:

- the duration of the signal T_{signal} is high compared to T_{pulse};
- the frequency bandwidth of e is within the interval $[-f_0/2; f_0/2]$.

Then its autocorrelation function is approximately

$$C_{S_{B,finite}}(\tau) \simeq \Lambda\left(\frac{\tau}{T_{signal}}\right)\left(\sum_{n\in\mathbb{Z}} |a_n|^2 e^{2i\pi n f_0 \tau}\right)$$
$$\times \left(\sum_{m\in\mathbb{Z}} |E(mf_{pulse})|^2 e^{2i\pi m f_{pulse}\tau}\right),$$

where $\Lambda(t)$ is the triangular function ($\Lambda(t) = 1 + t$ on $[-1; 0]$ and $\Lambda(t) = 1 - t$ on $[0,1]$ and zero outside of $[-1; 1]$).

Proof: The Fourier transform of $s_{B,finite}$ is (see previous paragraph)

$$S_{B,finite}(f) = \sum_{m\in\mathbb{Z}} E(mf_{pulse}) G_{f_0}(f - mf_{pulse}) * W(f)$$
$$= T_{signal}\left[\sum_{m\in\mathbb{Z}} E(mf_{pulse}) \times \sum_{m\in\mathbb{Z}} a_n \delta(f - mf_{pulse}\right.$$
$$\left. - nf_0) * \text{sin}_c(\pi T_{signal} f)\right](f)$$
$$= T_{signal} \sum_{n,m\in\mathbb{Z}} a_n E(mf_{pulse})$$
$$\times \text{sin}_c(\pi T_{signal}(f - mf_{pulse} - nf_0)). \quad \text{(A1)}$$

The Pancherel formula [Appel (2008), p. 263] applied to $s_{B,\text{finite}}(t)$ and $s^*_{B,\text{finite}}(t + \tau)$ gives that the autocorrelation function $C_{S_{B,\text{finite}}}(\tau)$ is the inverse Fourier transform of the squared spectral density $|FT(s_{B,\text{finite}})|^2(f)$ of the signal. Thus,

$$C_{S_{B,\text{finite}}}(\tau) = FT^{-1}\left(\left|\sum_{n,m\in\mathbb{Z}} a_n E(mf_{pulse}) \times T_{\text{signal}}\text{sinc}(\pi T_{\text{signal}}(f - mf_{pulse} - nf_0))\right|^2\right).$$

The two facts that the duration of the signal T_{signal} is high compared to T_{pulse} and that the bandwidth of e is within the interval $[-f_0/2; f_0/2]$ imply that for a particular f all but one term of this sum are very close to zero. Thus, we can say that

$$C_{S_{B,\text{finite}}}(\tau) \simeq FT^{-1}\left(\sum_{n,m\in\mathbb{Z}} |a_n|^2 |E(mf_{\text{pulse}})|^2 \times |T_{\text{signal}} \text{sinc}(\pi T_{\text{signal}}(f - mf_{\text{pulse}} - nf_0))|^2\right)$$

$$\simeq \sum_{n,m\in\mathbb{Z}} |a_n|^2 |E(mf_{\text{pulse}})|^2 \times FT^{-1}\left(|T_{\text{signal}} \text{sinc}(\pi T_{\text{signal}}(f - mf_{\text{pulse}} - nf_0))|^2\right)$$

$$\simeq \sum_{n,m\in\mathbb{Z}} |a_n|^2 |E(mf_{\text{pulse}})|^2 \times e^{2i\pi(nf_0 + mf_{\text{pulse}})\tau} \Lambda(\tau/T_{\text{signal}})$$

$$\simeq \Lambda(\tau/T_{\text{signal}})\left(\sum_{m\in\mathbb{Z}} |a_n|^2 e^{2i\pi nf_0\tau}\right) \times \left(\sum_{m\in\mathbb{Z}} |E(mf_{\text{pulse}})|^2 e^{2i\pi mf_{\text{pulse}}\tau}\right).$$

Remarks. The maximum of $\sum_{m\in\mathbb{Z}} |a_n|^2 e^{2i\pi nf_0\tau}$ is obtained when τ is an integer multiple of T_0 and the maximum of $\sum_{m\in\mathbb{Z}} |E(mf_{\text{pulse}})|^2 e^{2i\pi mf_{\text{pulse}}\tau}$ is obtained when τ is an integer multiple of T_{pulse}. Thus, in the case of a tonal signal where $T_{\text{pulse}} = k\,T_0$, we will have a maximum of $C_{S_{A,\text{finite}}}$ for $\tau = T_{\text{pulse}}$. In the case of a non-tonal signal (f_0/f_{pulse} is not an integer), we will have a maximum of $C_{S_{B,\text{finite}}}$ at the multiple of T_0 which is the nearest value to T_{pulse}. In this case the determination of T_{pulse} by autocorrelation has a bias.

A more detailed version of these proofs can be obtained by asking the authors.

Adam, O., Cazau, D., Gandilhon, N., Fabre, B., Laitman, J., and Reidenberg, J. (**2013**). "New acoustic model for humpback whale sound production," Appl. Acoust. **74**, 1182–1190.

Appel, W. (**2008**). *Mathématiques pour la Physique et les Physiciens* (H & K Editions, Paris).

Aroyan, J. L., McDonald, M. A., Webb, S. C., Hildebrand, J. A., Clark, D., Laitman, J. T., and Reidenberg, J. S. (**2000**). *Acoustic Models of Sound Production and Propagation* (Springer-Verlag, Berlin), Chap. 10, pp. 409–469.

Au, W., Popper, A., and Fay, R. (**2000**). *Hearing by Whales and Dolphins* (Springer, Berlin).

Balcazar, N., and Alosilla, G. (**2018**). (private communication).

Brown, J. C. (**2008**). "Mathematics of pulsed vocalizations with application to killer whale biphonation," J. Acoust. Soc. Am. **123**(5), 2875–2883.

Buchan, S., Hucke-Gaete, R., Rendell, L., and Stafford, K. (**2014**). "A new song recorded from blue whales in the Corcovado Gulf, southern Chile, and an acoustic link to the eastern tropical Pacific," Endang. Spec. Res. **23**, 241–252.

Cummings, W., and Thompson, P. (**1971**). "Underwater sounds from the blue whale, *Balaenoptera musculus*," J. Acoust. Soc. Am. **50**, 1193–1198.

Dziak, B., Haxel, J., Lau, T.-K., Heimlich, S., Caplan-Auerbach, J., Mellinger, D., Matsumoto, H., and Mate, B. (**2017**). "A pulsed-air model of blue whale b call vocalizations," Sci. Rep. **7**, 9122.

Eaton, J. W., Bateman, D., and Hauberg, S. (**2009**). *GNU Octave Version 3.0.1 Manual: A High-level Interactive Language for Numerical Computations*, CreateSpace Independent Publishing Platform, http://www.gnu.org/software/octave/doc/interpreter (Last viewed 17 September 2019).

Fant, G. (**1960**). *Acoustic Theory of Speech Production*. (Mouton, The Hague).

Flanagan, J. (**1965**). *Speech Analysis Synthesis and Perception* (Springer, Berlin).

Gavrilov, A., McCauley, R., Salgado-Kent, C., Tripovitch, J., and Wester, C. B. (**2011**). "Vocal characteristics of pygmy blue whales and their change over time," J. Acoust. Soc. Am. **130**(6), 3651–3660.

Glockner, D. A. (**1983**). "Determining the sex of humpback whales (*Megaptera novaeangliae*) in their natural environment," in *Communication and Behavior of Whales. AAAS Selected Symposium 76*, edited by R. Payne (Westview Press, Boulder, CO), pp. 447–464.

Glotin, H. (**2001**). "Dominant speaker detection based on voicing for adaptive audio-visual ASR robust to speech noise," in *ISCA Tutorial and Research Workshop (ITRW) on Adaptation Methods for Speech Recognition*.

Howard, D. M., and Angus, J. A. S. (**2006**). *Acoustics and Psychoacoustics* (Elsevier, Amsterdam).

McDonald, M., Hildebrand, J., and S. Mesnick (**2009**). "Worldwide decline in tonal frequencies of blue whale songs," Endang. Spec. Res. **9**, 13–21.

McDonald, M., Mesnik, S., and Hildebrand, J. (**2006**). "Biogeographic characterisation of blue whale song worldwide: Using song to identify populations," J. Cetacean Res. Manage. **8**(1) 55–65.

Mercado, E. III, Schneider, J. N., Pack, A. A., and Herman, L. M. (**2010**). "Sound production by singing humpback whales," J. Acoust. Soc. Am. **127**(4), 2678–2691.

Miller, B. S., Collins, K., Barlow, J., Calderan, S., Leaper, R., McDonald, M., Ensor, P., Olson, P., Olavarria, C., and Double, M. (**2014**). "Blue whale vocalizations recorded around New Zealand: 1964–2013," J. Acoust. Soc. Am. **135**(3), 1616–1623.

Miralles, R., Lara, G., Antonio, E., and Rodriguez, A. (**2012**). "The pulsed to tonal strength parameter and its importance in characterizing and classifying beluga whale sounds," J. Acoust. Soc. Am. **131**(3), 2173–2179.

Murray, S. O., Mercado, E. III, and Roitblat, H. L. (**1998**). "Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations," J. Acoust. Soc. Am. **104**(3), 1679–1688.

Nowicki, S., and Capranica, R. R. (**1986**). "Bilateral syringeal interaction in the production of an oscine bird sound," Science **231**, 1297–1299.

Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. (**2007**). "Behavioral context of call production by eastern north Pacific blue whales," Mar. Ecol. Prog. Ser. **330**, 269–284.

Patris, J., Malige, F., and Glotin, H. (**2017**). "Construction et mise en place d'un système fixe d'enregistrement à large bande pour les cétacés "Bombyx 2" Isla de Chañaral, été austral 2017," Technical Report 2017-03.

Rankin, S., Oswald, J., Barlow, J., and Lammers, M. (**2007**). "Patterned burst-pulse vocalizations of the northern right whale dolphin, *Lissodelphis borealis*," J. Acoust. Soc. Am. **121**(2), 1213–1218.

Reidenberg, J. (**2017**). "Terrestrial, semiaquatic, and fully aquatic mammal sound production mechanisms," Acoust. Today **13**(2), 35–43.

Richardson, W. J., Charles R., Greene, J., Malme, C. I., Thomson, D. H., Moore, S. E., and Wiirsig, B. (**1995**). *Marine Mammals and Noise* (Academic Press, New York).

Simard, P., Mann, D. A., and Gowans, S. (**2008**). "Burst-pulse sounds recorded from white-beaked dolphins (*Lagenorhynchus albirostris*)," Aquat. Mamm. **34**(4) 464–470.

Stafford, K., Chapp, E., and Bohnenstiel, D. (**2011**). "Seasonal detection of three types of 'pygmy' blue whale calls in the Indian Ocean," Marine Mamm. Sci. **27**(4), 828–840.

Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999**). "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific," J. Acoust. Soc. Am. **106**(6), 3687–3698.

Watkins, W. (**1968**). *The Harmonic Interval Fact or Artifact in Spectral Analysis of Pulse Train* (Pergamon Press, Oxford), Vol. 2.

Wise, J., Caprio, J., and Parks, T. (**1976**). "Maximum likelihood pitch estimation," Trans. Acoust. Speech Sign. Process. **24**, 418–423.

2154   J. Acoust. Soc. Am. **146** (4), October 2019

Patris *et al.*