Research Article

Sparse Representation Based on Tunable Q-Factor Wavelet Transform for Whale Click and Whistle Extraction

Hailan Chen, Jiaquan Yan, Naveed Ur Rehman Junejo, Jie Qi, and Haixin Sun

1 School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
2 School of Science, Jimei University, Xiamen 361021, China

Correspondence should be addressed to Haixin Sun; hxsun@xmu.edu.cn

Received 25 May 2018; Revised 27 August 2018; Accepted 24 September 2018; Published 4 November 2018

Academic Editor: Cristina Castejón

Whale sounds may mix several elements including whistle, click, and creak in the same vocalization, which may overlap in time and frequency, so it leads to conventional signal separation techniques challenging to be applied for the signal extraction. Unlike conventional signal separation techniques which are based on the frequency bands, such as WT and EMD, tunable-Q wavelet transform (TQWT) can separate the objected signal into particular components with different structures according to its oscillation property and eliminate in-band noise using the basis pursuit method. Considering the characteristics of oscillatory and transient impulse, we propose a novel signal separation method for whale whistle and click extraction. The proposed method is performed by the following two steps: first, TQWT is used to construct the dictionary for sparse representation. Secondly, the whale click and whistle construction are performed by designing the basis pursuit denoising (BPD) algorithm. The proposed method has been compared with one of the popular signal decomposition techniques, i.e., the EMD method. The experimental results show that the proposed method has a better performance of click and whistle signal separation in comparison with the EMD algorithm.

1. Introduction

Whale sounds are combined of several elements (whistle, regular click, and rapid-click buzzes (creak)) in the same vocalization [1]. In general, whales utilize frequency-modulated pure tones (whistles) to communicate with each other. Meanwhile, they emit transient impulses (clicks) to echolocate the targets and explore the environment. Besides, they vocalize creak voice when they are in danger or emergency. However, most of the existing methods in the field of whale signal processing have been only analyzed for the single whistle or click signals, such as in [2–10].

Rather than extracting useful information from the multicomponent signals for further processing, some methods directly extract contours from whale whistle using image processing techniques, as described in [2–5]. Besides, for the automated analysis of whale click, various methods have been presented based on the statistical computation of whale click spectrogram through different transforms in [6–10] with only presence of click. To further cognition of whale communication pattern and echolocation pattern, we need to separately extract whistles and clicks from the composite signal. Abundant methods have been presented to decompose multicomponent signals such as blind source separation [11], dual-tree complex wavelet transform [12], wavelet denoising, empirical mode decomposition (EMD) [13], ensemble empirical mode decomposition (EEMD), multiwavelet packet [14], and independent component analysis (ICA) [15–18]. The above methods are managed to decompose the objected signals in the frequency domain. However, each component of the whale sounds may occupy the same frequency band and overlap in the frequency domain. Thus, the above methods cannot exactly extract each component from the whale sounds.

Unlike frequency-based methods, a sparse signal representation method using the tunable Q factor wavelet
transform (TQWT) can be used to decompose multicomponent signal according to the oscillatory behavior, where the oscillation is defined and reflected by the Q factor. Sparse signal representation has been applied to signal separation. An approach using biorthogonal RADWTs for signal separation was presented in [19]. A method using a signal sparse representation for signal separation has been proposed to address the mitigation of wind turbine clutter (WTC) in weather radar data [20]. Since the whistle and click signals have oscillatory and transient impulse characteristics, respectively, we proposed a sparse signal representation method with TQWT to extract click components with low Q factor and whistle components with high Q factor, respectively. The method is verified by the oceanic audio recording, and the results show effective extraction of click and whistle components from whale vocalization. The main contribution of this paper lies in (1) a new method that extracts click and whistle effectively and efficiently from multicomponents sound, (2) the selection of appropriate algorithm parameters in the actual application, and (3) the proposed method that can also be used for other mammals to extract interesting component from the composite signal.

The rest of the paper is organized as follows: in Section 2, the TQWT algorithm and the basis pursuit denoising method are introduced. Section 3 compares the capabilities of the proposed approach and the EMD algorithm using click and whistle components extraction from a recorded whale composite sound signal. Section 4 discusses the conclusion.

2. Methodology

The proposed approach consists of two parts: TQWT and BPD. TQWT offers an excellent flexibility for representing a signal of different temporal and spectral characteristics, achieved by tuning the transform parameters. TQWT is a discrete-time wavelet transform (DWT) [21] which provides a suitable tool for the analysis of oscillatory and nonoscillatory components of a signal. Besides, BPD is applied to obtain a sparse representation of each component which has different oscillatory behaviors. A brief introduction of the TQWT, its oscillation characteristic, and the transform operation are discussed in the following section.

2.1. TQWT. Our aim is to separate click and whistle components from the composite signal. From the perspective of the oscillation nature, the click components are composed of instantaneous pulses which have low oscillation property, and the whistle components are constituted of multiharmonics which have high oscillation property. To deal with such signals, Selesnick et al. developed the tunable Q factor wavelet transform (TQWT) which the Q factor is flexibly tunable [22]. TQWT can be used to decompose the objected signal into high-oscillatory, low-oscillatory, and residual components according to the values of high Q and low Q factors [23]. The Q factor reflects the oscillatory properties of one signal. The Q factor is defined as follows [24]:

\[
Q = \frac{f_c}{\text{BW}},
\]

where \( f_c \) is the center frequency and BW is the bandwidth. The signal oscillatory property can be described with the Q factor. As shown in Figures 1(a)–1(d), it is evident that a higher Q signal has a higher oscillatory intensity in the time domain and a higher degree of frequency aggregation in the frequency domain at the same time, and vice versa [24]. Hence, the difference between the low and high Q wavelet functions highlights the idea of oscillation of a signal, which is exploited for the component extraction problem discussed in this paper.

The TQWT algorithm which decomposes an \( N \)-point discrete-time signal into \( J \)-level subbands is demonstrated in Figure 2. The structure of TQWT employs two channel analysis and synthesis filter banks. The implementation of the analysis filter banks is performed on its low-pass channel iteratively and then further processed by the low- and high-pass scaling, and \( \alpha \) and \( \beta \) are the corresponding scaling parameters [25]. The synthesis filter bands execute the same steps. For each level, two-channel filters are composed of low- and high-pass filter, and the corresponding frequency responses are defined in the following equations [23]:

\[
H_1(\omega) = \begin{cases} 
1, & |\omega| \leq (1 - \beta)\pi, \\
\frac{\theta((\omega + (\beta - 1)\pi)\alpha + \beta - 1)}{(\alpha + \beta - 1)}, & (1 - \beta)\pi < |\omega| < \alpha\pi, \\
0, & \alpha\pi \leq |\omega| \leq \pi,
\end{cases}
\]

\[
H_\text{h}(\omega) = \begin{cases} 
0, & |\omega| \leq (1 - \beta)\pi, \\
\frac{\theta((\alpha\pi - \omega)\alpha + \beta - 1)}{(\alpha + \beta - 1)}, & (1 - \beta)\pi < |\omega| < \alpha\pi, \\
1, & \alpha\pi \leq |\omega| \leq \pi,
\end{cases}
\]

where the parameters must satisfy \( 0 < \alpha < 1, \ 0 < \beta \leq 1, \ \alpha + \beta > 1, \) and \( \theta(\omega) = 0.5(1 + \cos\omega)\sqrt{2 - \cos\omega}, \ |\omega| \leq \pi \).

The most significant parameters of the TQWT algorithm are Q factor, redundancy factor \( r \), and decomposition level \( J \). The Q factor describes the degree of signal oscillation. For a high Q factor, the wavelets have more dense oscillatory cycles, which are suitable for the extraction of oscillatory components. Meanwhile, for a low Q factor, the wavelets consist of nonoscillatory elements, which are fit for the extraction of the transient components. The redundancy factor \( r \) controls the overlapping rate among the frequency responses of the adjacent wavelets. Increasing the \( r \) value in the case of a fixed Q value enhances the overlapping rate of frequency responses and the computational cost. Note that \( r \) must be greater than 1, and \( r \geq 3 \) is recommended for the perfect reconstruction and sparsity. The value of the decomposition level \( J \) affects the frequency coverage of the wavelets. Greater \( J \) value makes the wavelets cover a wider
frequency range which even reach 0Hz. The value of $J$ should be selected as large as possible to include the lower frequency to the utmost extent. The maximum number of decomposing level $J_{\text{max}}$ is set to

$$J_{\text{max}} = \left\lfloor \log\left(\frac{\beta N}{8}\right)/\log\left(\frac{1}{\alpha}\right) \right\rfloor \frac{\log\left(\frac{N}{4} (Q + 1)\right)}{\log\left(\frac{(Q + 1) r}{(Q + 1) r - 2}\right)} ,$$

where $N$ is the length of the input signal $x(n)$. 

Figure 1: Wavelet waveform and the corresponding frequency response spectrum with different parameters: (a) waveform with $Q = 1$ and $r = 3$, (b) spectrum with $Q = 1$ and $r = 3$, (c) waveform with $Q = 3$ and $r = 3$, and (d) spectrum with $Q = 3$ and $r = 3$.

Figure 2: The filter banks of TQWT. (a) The analysis filter banks. (b) The synthetic filter banks.
In the practical applications, we calculated the value of $f_{\text{max}}$ according to Equation (3). Moreover, the computational cost of the TQWT algorithm is $O(3rN \log_3(4N/(Q + 1)))$ [23], and we discuss the computational complexity with the change of parameters. We can properly get the conclusion that the computational complexity increases with the enlargement of $Q$ and $J$ values, but increasing the Q value reduces the computational cost in the case of $J = J_{\text{max}}$. Although sometimes we must select big enough $Q$ value to match the high oscillatory intensity of the object signal and set $J = J_{\text{max}}$, but the computational cost is not huge according to the above discussion.

In essence, the TQWT algorithm can provide one set of overcomplete basis to estimate high- and low-resonance components. We need to select three parameters $Q$, $r$, and $J$ to establish the basis. The selection guides of the above parameters are mentioned in Table 1 [26].

### 2.2. Basis Pursuit Denoising (BPD) with TQWT

Sparsity can be used in signal processing problems. By considering an observed signal $y(n)$ which has been corrupted by additive noise, the received signal can be represented as

$$y(n) = x(n) + i(n),$$

where $x(n)$ is the useful signal and $i(n)$ is the additive noise. The problem is to estimate $x(n)$ which has a sparse representation from the observed signal $y(n)$. The technique of BPD [27] provides an excellent platform for the optimization of the sparse representation of the signal $x(n)$. To obtain a sparse representation of the signal $x(n)$ concerning TQWT, we construct the objection function based on BPD technique as

$$\arg\min_{\omega} \left\{ \|y - \text{TQWT}^{-1}(\omega)\|_2^2 + \lambda \|\omega\|_1 \right\},$$

where $\| \cdot \|_1$ and $\| \cdot \|_2$ denote $L_1$ and $L_2$ norms, respectively. The coefficient $\omega = [\omega^{(1)}, \ldots, \omega^{(J+1)}]$ is the wavelet coefficient computed by TQWT. $\omega^{(j)}$ denotes the wavelet coefficient of the subband $j$, $\eta$ is a “regularization” parameter, $\odot$ denotes the Hadamard (elementwise) multiplication, and $\lambda = (\lambda_1, \ldots, \lambda_{J+1})$ is the compensation vector.

Several different algorithms have been developed for solving basis pursuit (BP) and BPD problems, such as iterative shrinkage/threshold algorithm (ISTA) [28, 29], fast ISTA (FISTA) [30], and split variable augmented Lagrangian shrinkage algorithm (SALSA) [31, 32]. The main idea of SALSA method is summarized in Algorithm 1, where $\mu$ represents the penalty parameter which affects the algorithm convergence speed and $P$ is the iteration number. Considering the length of the article, the details of SALSA algorithm can be referred to [31].

### 3. Experimental Results and Analysis

In the experiment, we adopted a real oceanic audio recording of a Beluga whale which consists of chick and whistle components. This recording was acquired at the Oceanographic Valencia, using the instruments such as a computer with a Roland (Edirol) FA-101 sound acquisition system, a Brul and Kjaer 8103 hydrophone, and a Brul and Kjaer 2692 Nexus amplifier, as described in [1]. The waveform, spectrum, and time-frequency signature representation of the raw signal are plotted in Figures 3, 4(a), and 5(a), respectively. In Figure 3, there are 8192 samples in total under the sampling frequency $f_s = 96$ kHz for the test. It is evident that the wave exists several transient impulses which denote click components. Meanwhile, we can observe from the subfigure of Figure 3 at $t = 0.0601 \sim 0.0611$ s that they have oscillatory components. Both click components and whistle components are clearly visualized in the spectrum as demonstrated in Figure 5(a): clicks are broadband components in the whole frequency domain. On the contrast, whistles are impulsive components which are distributed from 8 kHz to 26 kHz. We can clearly see from Figure 4(a) that whistles and click components intersect in the time-frequency signature. Meanwhile, whistles are line spectrum in the frequency axis, and clicks are also discrete distribution in the time axis.
3.1. Case Study of Whale Click Extraction. According to the nonoscillatory characteristic nature of click signal and the selection criteria of the TQWT approach, we set $Q = 1$, $r = 3$, and $J = 17$ for click extraction. After obtaining the wavelet coefficients by the TQWT method, the BPD algorithm is performed to reconstruct the click signal. To solve the basis...
Figure 5: Comparison of the effectiveness of whistle and click components separation by the proposed method and the EMD algorithm. (a) The raw signal. (b) The proposed method. (c) The EMD algorithm.
pursuit problem, we used the SALSA algorithm with parameters $\mu = 2$, $\eta = 0.06$, and $P = 100$. After implementing two above-mentioned methods, we obtain the sparse representation of extracted click components. The time-frequency signature of the extracted click components is shown in Figure 4(c). Compared with Figure 4(a), it is obvious that all the click components of Figures 4(a) and 4(c) are spread at the same position in the time domain, and their shapes are also identical. In addition, the resolution of Figure 4(c) is much purer than Figure 4(a), because the BPD algorithm can eliminate the noise. However, we have marked weak and strong waveformsof the click signal with ellipse and rectangle, respectively, as seen in Figure 6. As compared with the two subfigures in Figure 6, the weak click signal has low energy, which is almost as harmonics, so, it cannot be extracted effectively. Further, to verify the effectiveness of the click extraction, the waveform and the spectra of the raw and extracted signals are drawn in Figures 5(b) and 6, respectively. We concluded that the performance of extracting the click is quite remarkable, and the proposed method has a good effect on noise reduction.

3.2. Case Study of Whale Whistle Extraction. Similarly, considering the oscillatory nature of whistle signal, we set $Q = 320$, $r = 3$, and $f = 891$. To obtain whistle component sparse representation concerning TQWT, the BPD approach based on the SALSA algorithm is utilized with parameters $\mu = 2$, $\eta = 0.045$, and $P = 100$.

The time-frequency signature of the whistle components extracted by the proposed method is drawn in Figure 4(b). We can find that whistle components of Figures 4(a) and 4(b) are distributed at the same spot and their shapes are horizontal lines. We compare the effect of whistle and click signal separation by the proposed method and the EMD algorithm in Figures 5(b) and 5(c). It is obvious that the spectra of the whistle and click components separated by the proposed method are identical to the spectra of the raw signal. Furthermore, we can draw the conclusions that the proposed method can separate the whistle and click components from the raw signal effectively and the proposed method has the function of noise reduction to some extent.

In this paper, the EMD algorithm was selected as a comparison due to its advantage of analyzing the non-linear and nonstationary signals. EMD method can separate the objected signal into different intrinsic mode functions (IMFs). As exhibited in Figure 5(b), the first three IMFs are drawn to visualize its spectra where most of the harmonic waves can be extracted by the EMD algorithm, which represent the whistle components. However, click components cannot be extorted by EMD method. In addition, the noise components with the relative higher amplitude are distributed all over the frequency domain. In conclusion, the proposed method has the better performance of signal separation of the whistle and click components from the raw signal and the stronger ability of noise suppression.

4. Conclusion

In this paper, we have proposed a sparse representation with TQWT method to extort the click and whistle components from the raw whale signal according to the oscillatory behavior of these two components. Considering the disparate oscillatory intensity of click and whistle signals, the TQWT algorithm with high Q factor was operated to extract whistle component. On the other hand, click extraction is realized under the condition of low Q factor when we perform TQWT method. Besides, BPD is used as an alternative to LTI filters for noise reduction. The
proposed method and the EMD algorithm are applied to the same real whale vocalization. Our proposed method appears to be superior to EMD for the extraction of the precise frequency variation of the whistle. In conclusion, the proposed method has the prominent performance of extracting the click and whistle components and removing noise.

Data Availability
The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest
The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (Nos. 61671394 and 61471308), the Young and Middle-aged Teachers Educational Scientific Research Project of Fujian Province in 2016 (No. JAT160268), the Fundamental Research Funds for the Central Universities (No. 20720170044), and the Science and Technology Program of Shenzhen (JSGG20170141090428464).

References


