Multimodal sparse time-frequency representation for underwater acoustic signals

Yongchun Miao, Jianghui Li, *Member, IEEE* and Haixin Sun, *Member, IEEE*

Abstract

Multiple features can be extracted from time-frequency representation (TFR) of signals for the purpose of acoustic event detection. However, many underwater acoustic signals are formed by multiple events (impulsive and tonal), which generates difficulty on the high-resolution TFR for each component. For the characterization of such different events, we propose an anisotropic chirplet transform to achieve the TFR with high energy concentration. Such transform applies a time-frequency-varying Gaussian window to compensate the energy of each component while suppressing unwanted noise. Using a set of directional chirplet ridges from the obtained TFR, a structure-split-merge algorithm is designed to reconstruct a multimodal sparse representation, which provides instantaneous frequency and time features. Specifically, a pulsed-to-tonal ratio, based on these features, is computed to distinguish two acoustic signals. The presented method is validated using shallow water experimental underwater acoustic communication signals, and large sequences of harmonics and pulsed bursts from common whales.

Index Terms

Anisotropic chirplet transform (ACT), multimodal sparse representation (MSR), pulsed-to-tonal ratio (PTR), time-frequency representation (TFR), underwater acoustic (UWA) signals.

List of symbols

s	Signal	s_i	Each mode of a signal
t_i	Time point of i-th mode	ω_i	Frequency point of i-th mode
$\hat{t}_i$	Instantaneous time (IT)	$\hat{\omega}_i$	Instantaneous frequency (IF)
S	Time-frequency transform	c	Chirp rate
δ	Dirac delta function	$\hat{\omega}'$	The second derivative of a phase
ϑ	Instantaneous rotating operators (IROs)	$\arctan\vartheta$	Instantaneous rotating angles (IRAs)
g_c^*	Conjugation kernel of chirplet transfrom (CT)	h_{σ_t}	Analysis window of CT
σ_t	Time scale	σ_f	Frequency scale
L	Gaussian window width	$\hat{h}$	TF-varying Gaussian window
λ	Anisotropic operator	μ	Height-to-width ratio of window
$\sigma_{\bar{\varepsilon}}^2$	Variance of the smoothed noise	$\sum_S$	Energy sum in sub-contours of the spectrogram S
ρ	Observed energy function	$\hat{S}$	Multimodal sparse representation (MSR)

Yongchun Miao and Haixin Sun are with the School of Informatics, Xiamen University, Xiamen, Fujian, 361005, China (e-mail: ycmiao@stu.xmu.edu.cn and hxsun@xmu.edu.cn)

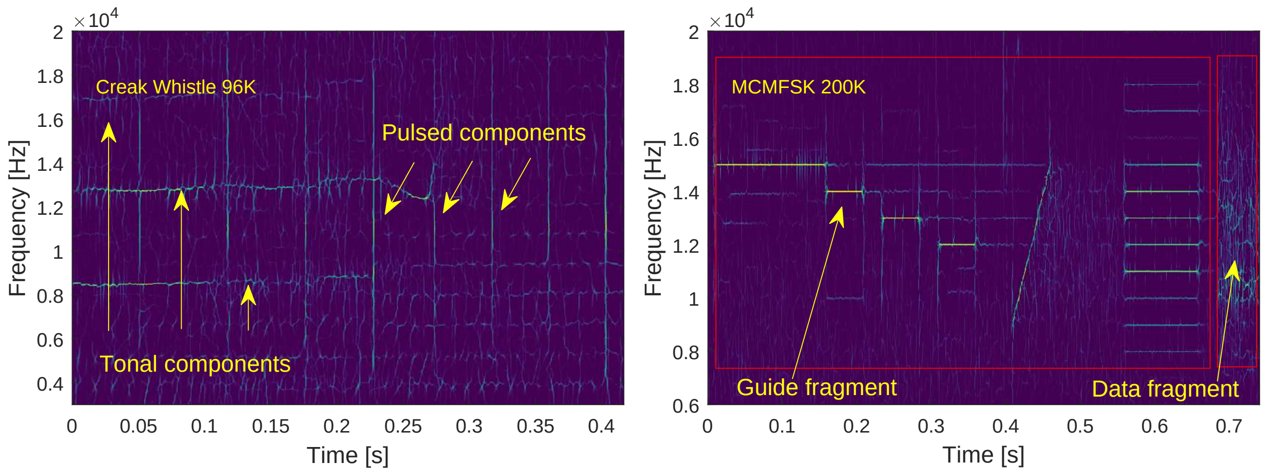
Jianghui Li is with the Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, U.K. (e-mail: J.Li@soton.ac.uk)

Corresponding author: Haixin Sun and Jianghui Li

Multimodal sparse time-frequency representation for underwater acoustic signals

I. INTRODUCTION

THE detection of different acoustic events is an important task in distinguishing between underwater acoustic (UWA) signals, e.g., marine mammal sounds, UWA communication signals, shipping noise and sonar pulses [1]–[8]. However, the characteristic analysis is usually difficult, because many UWA signals display low signal-to-noise ratio (SNR) or multiple modes. An instance of a whale signal (Fig. 1(a)), ranging from short pulsed transients to long tones, has various mode structures. To analyze such UWA signals, multiple features are used to characterize events of interest and merged together to solve the detection problem, and the time-frequency (TF) analysis is an efficient technique for the representation of multiple features [2], [5], [9]–[12].



(a) TFR of a creak whistle sound.

(b) TFR of a MCMFSK signal.

Fig. 1. Time-frequency representation (TFR) of sound generated by a whale and a multi-carrier multiple frequency-shift keying (MCMFSK) modulated signal. (a) TFR of a creak whistle sound; (b) TFR of a MCMFSK signal. The guide fragment is a pilot part of a modulation signal.

The main goal of the TF analysis is to determine the energy distribution along the frequency axis at each time instant. When the analysed UWA signal is multiple components (modes), defined by

$$s(t) = \sum_{i=1}^{N} s_i(t) + n(t) \tag{1}$$

with each mode $s_i(t) = a(t)e^{j\phi(t)}$, where $n(t)$ is a noise representing any undesirable components, and a representation of each element's energy concentration should be expected. The analytical signal $s(t)$ is generated by the Hilbert transform $\boldsymbol{H}$, i.e., $s(t) = x(t) + jH[x(t)]$. The majority energy of each component $s_i(t)$ is concentrated on the TF location (t_i, ω_i). The instantaneous frequency (IF) [13], [14] and instantaneous time (IT) [14] provide a measurement index of the energy concentration for the $s_i(t)$ defined by the time derivative of the phase $\hat{\omega}_i(t,\omega) = \frac{1}{2\pi}\frac{d\phi}{dt}$, and the subtraction of current time and the frequency derivative of the phase $\hat{t}_i(t,\omega) = t - \frac{1}{2\pi}\frac{d\phi}{d\omega}$ [14], respectively. The transformation $(t_i, \omega_i) \rightarrow (\hat{t}_i(t,\omega), \hat{\omega}_i(t,\omega))$ reveals multiple features of acoustic events. For a single tone of frequency ω_k, the (t_i, ω_i) pairs are transformed into several lines $(\hat{t}_i(t,\omega), \omega_k)$, such as a multi-carrier multiple frequency-shift keying (MCMFSK) signal as shown in Fig. 1(b); for a "click" whale signal as shown in Fig. 1(a), Dirac delta function $\delta(t - t_k)$ localized at time t_k, all energy points are transformed into the vertical lines $(t_k, \hat{\omega}_i(t,\omega))$.

The classical TF analysis includes short time Fourier transform (STFT) [15], continuous wavelet transform (CWT) [16]–[18], Wigner-Ville distribution (WVD) [10], [19], [20] and chirplet transform (CT) [21]–[24]. Energy concentration of the STFT and the CWT is limited by the Heisenberg uncertainty principle [15], [16], [21], [22]. The WVD [19], [20] can obtain good TF resolution, but when multicomponent signals are considered, the undesired cross terms appear in addition to signal components referred to as the auto terms, reducing the readability of the time-frequency representation (TFR). The CT offers more flexibility for the TFR than the STFT with an unmodulated basis [21], [22], and becomes a natural tool to analyse the chirp signals. However, these methods obtain the TFR with low-resolution energy concentration.

Currently, there are often two general strategies for the enhancement of the energy concentration: 1) Reassignment techniques aim to sharpen the TF representation by assigning the local energy while removing most of the interference [2], [14], [25],

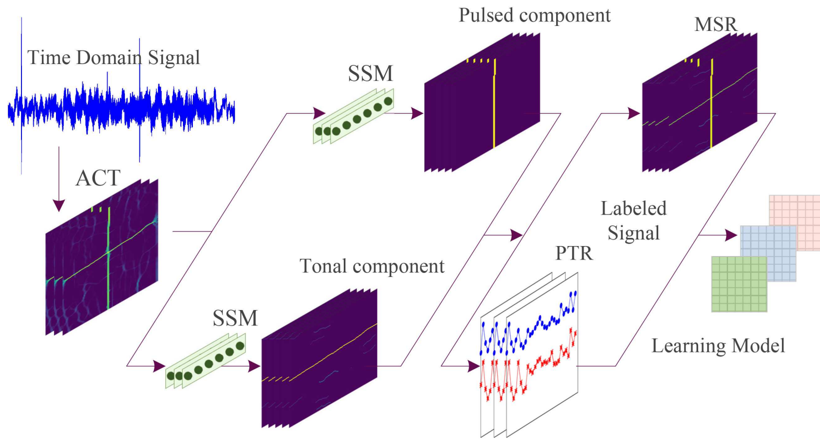


Fig. 2. The main framework of the proposed methods. An anisotropic chirplet transform (ACT) provides the TFR of signal. Using the chirplet ridges, SSM decomposes and extracts tonal and pulsed features from the TFR. Further, we obtain a pulsed-to-tonal ratio (PTR) and the multimodal sparse representation (MSR) to discriminate UWA signals.

[26]. Numerous reassigned TF representations have been proposed such as second-order synchrosqueezing transform [14] and Fourier synchrosqueezing transform (FSST) [25]. The reconstruction of TFR depends on the maximum of the spectrogram. Since the high noise induces false maxima (maxima outside of the auto terms) in the TF plane, the methods are sensitive to noise. The reassignment processing also possesses high computational cost. 2) Several energy concentration measuring methods [1], [4], [27] have been presented for the optimization of window width by applying special algorithms. Similar to the S-transform [27], an adjustment of the window size depends on local signal characteristics. The search of optimal values usually suffers from heavy computational complexity.

The chirplet path fusion [28], cubic phase function [29], [30], and adaptive fractional spectrogram [13], [31], [32] have low computational cost to analyze noisy signals with closely-spaced chirps. To enhance the energy concentration of overlapping signals under high level of noise, a synchro-compensating chirplet transform [23] utilizes different chirp rates for different signal components. Since the condition $\hat{\omega}'(t,\omega) < \infty$ or $\hat{\omega}'(t,\omega) < \varepsilon$ in impulsive components of UWA signals no longer hold, the TFR obtained by the Gaussian window [23] is still blurred at transient points. In fact, the standard deviation of the one-dimensional Gaussian window [13], [16], [23] is proportional to the reciprocal of frequency. No matter what kind of signals are analysed, the width of the Gaussian window monotonically decreases as the frequency increases. Considering the time-scale and frequency-scale measurement simultaneously, the two-dimensional Gaussian window can be applied iteratively in order to improve the estimation reliability. Moreover, detecting TF ridges [23], [33] seems to be effective for the analysis of UWA signals consisting of several modes with similar magnitude overlapping in a significant portion of the TF plane.

In this paper, we focus on a multimodal sparse representation of signals derived from the ridges of anisotropic chirplet transform (separation ability). The main contributions of this work are shown as follows (Fig. 2).

1) For UWA signals, an anisotropic chirplet transform (ACT) is proposed to achieve a high-resolution TFR. A time-frequency-varying Gaussian window of the ACT is designed for the enhancement of energy concentration and allows an adjustment between complexity and TF resolution. Besides, since the optimal anisotropic operator is selected by maximizing the SNR, the ACT is less sensitive to noise.
2) Using second-order directional derivatives, the chirplet ridges of multimodal structures are defined in terms of points $(\hat{t}_i(t,\omega), \hat{\omega}_i(t,\omega))$ and extracted from the TFR obtained by the ACT. A structure-split-merge (SSM) algorithm is proposed to split the ridges to different TF structures and merge these structures into the sparse representation of tonal and pulsed components.
3) An improved pulsed-to-tonal ratio (PTR) utilizes the multimodal sparse representation of different components to discriminate similar UWA signals, e.g., marine biological signals.

The rest of this paper is organized as follows. Section II introduces the basic principle of the ACT, and the TF-varying window width. Details of the structure-split-merge (SSM) algorithm based on ACT are provided in Section III. Section IV applies experimental data to prove the efficiency of the presented algorithm, and illustrate the improvement in readability and detection of UWA signals. Finally, conclusions are drawn in Section V.

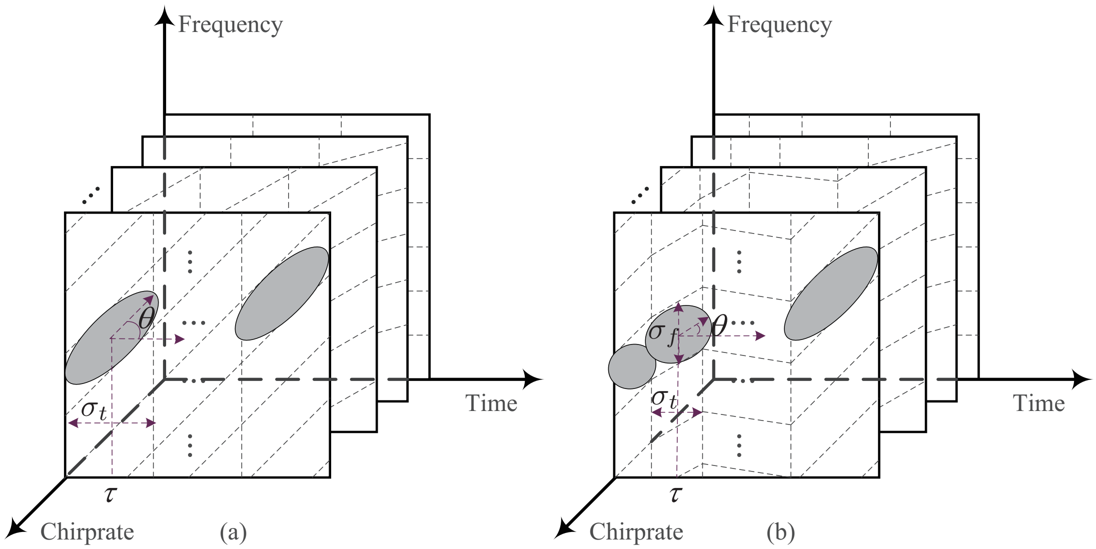


Fig. 3. Volumetric family of "Time-Frequency-Chirprate" transform. (a) Classical CT. (b) ACT.

II. ANISOTROPIC CHIRPLET TRANSFORM

A. TFR of multimodal signal

The chirplet transform (CT) [21] $S(t,\omega)$ can be formulated as the inner product of the signal $s(t)$ with chirplets:

$$\begin{aligned} S(t,\omega) &= \int_{-\infty}^{\infty} s(\tau) g_c^*(\tau - t, \omega) \mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} s(\tau) h_{\sigma_t}(\tau - t) \mathrm{e}^{j\frac{c}{2}(\tau-t)^2} \mathrm{e}^{-j2\pi\omega\tau} \mathrm{d}\tau, \end{aligned} \tag{2}$$

where t, ω and c are the time, frequency and chirp rate indices, respectively. The variable σ_t defines the temporal resolution of analysis $\triangle t$. The chirplet transform is a linear and unitary transform, rotating an angle θ with a single chirp rate c_0 as shown in Fig. 3(a). The angle is computed by a function of $\arctan(2T_s/F_s \times c_0)$, where T_s and F_s are the sampling time and sampling frequency, respectively [34]. If we utilizes different chirp rate $c(t)$ and rotation angle continuously and repeat the process uncountably for a number of times, a different volumetric family of ACT is expected to be generated (Fig. 3(b)). The analysis window $h_{\sigma_t}(t)$ of the CT is represented by the standard deviation σ_t centred at time $t = \tau$. Unlike the CT, the ACT possesses different window scales corresponding to the chirp rates. This implies that large values of σ_t and σ_f should be adopted if the instantaneous frequency (IF) of the signal varies smoothly, while small values of σ_t and σ_f should be adopted if the IF is fast-varying.

The main goal of the ACT is to apply different chirp rates for different signal components. For such purpose, a dominant instantaneous chirp rate coefficient at the interval t is defined as

$$\alpha(t,c)) = \max\left\{|S(t,\omega)|\right\}. \tag{3}$$

Maximum of $\alpha(t,c)$ generates a robust estimate of appropriate instantaneous rotating angles (IRAs) as

$$\vartheta(t) = \arg\max_c\{\alpha(t,c)\}, \tag{4}$$

where instantaneous rotating operators (IROs) $\vartheta(t)$ define IRAs $\arctan\vartheta(t)$ in the TF plane. The estimated chirp rates $c(t)$ is then determined using $\tan(\vartheta(t)) \times F_s/(2T_s)$.

For a chirp $s(t) = a(t)e^{j\int\varphi(\tau)\mathrm{d}\tau}$ signal, the transformation to all angles $\vartheta(t)$ is generated as

$$\varrho(t,\omega) = \frac{1}{\sigma_t}\int_{-\infty}^{\infty} (\tau - t) s(\tau) g_{c(t)}^*(\tau - t, \omega) \mathrm{d}\tau. \tag{5}$$

The set of points $(\hat{t}_i(t,\omega), \hat{\omega}_i(t,\omega))$ is defined by stationary phase positions $\hat{t}_i(t,\omega) = 0$ and $\hat{\omega}_i(t,\omega) = 0$. We obtain the points in a closed form

$$\begin{aligned} \hat{\omega}_i(t,\omega) &= \partial_t \Im \ln S = \omega + \frac{1}{\sigma_t}\Im\left\{\frac{\varrho}{S}e^{j\vartheta(t)}\right\}, \\ \hat{t}_i(t,\omega) &= t - \partial_\omega \Im \ln S = t + \sigma_t \Re\left\{\frac{\varrho}{S}e^{j\vartheta(t)}\right\}. \end{aligned} \tag{6}$$

The $\int \varphi(\tau)\mathrm{d}\tau$ of the chirp signal can be expanded by Taylor expansion $\hat{\omega}_i + \hat{\omega}_i'(\tau - t)$. The spectrogram of CT in the IF $\hat{\omega}_i$ is equivalent to

$$\begin{aligned} S(t, \hat{\omega}_i) &= \int_{-\infty}^{\infty} a(\tau)h(\tau - t)\mathrm{e}^{j\pi\left[\hat{\omega}_i' + c(t)\right](\tau - t)^2}\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} z(\tau)\mathrm{e}^{j\psi(\tau)}\mathrm{d}\tau, \end{aligned} \tag{7}$$

where both integrable $z(\tau) > 0$ and $\psi(\tau)$ are C^1, and $\hat{\omega}_i' + c(t) \neq 0$, the case $\hat{\omega}_i' + c(t) = 0$, resulting in a typical Fourier transform.

Assuming that $z(\tau) > 0$ is slowly-varying as compared to the oscillations controlled by the phase $\psi(\tau)$, the spectrogram can be obtained by the stationary phase approximation [29], [35]. In this situation, negative and positive values of the integrand tend to cancel each other, with the result that the main contribution to $S(t, \omega)$ only stems from the proximity of points where the phase is stationary. For the stationary point at a time τ_0, the derivative of the phase $\psi(\tau)$ will be zero:

$$\psi'(\tau_0) = 2\pi\left[\hat{\omega}_i' + c(t)\right](\tau_0 - t) = 0. \tag{8}$$

In the view of $\psi''(\tau_0) = 2\pi\left[\hat{\omega}_i' + c(t)\right]$, the phase of a linear chirp signal can be considered as a Taylor expansion $\psi(\tau) = \psi(\tau_0) + \frac{\psi''(\tau_0)}{2}(\tau - \tau_0)^2$ approximately. Then, the spectrogram (7) can be written as

$$S(t, \hat{\omega}_i) = z(\tau_0)\int_{-\infty}^{\infty} \mathrm{e}^{j\frac{\psi''(\tau_0)}{2}(\tau - \tau_0)^2}\mathrm{d}\tau. \tag{9}$$

Letting the change of variables

$$v^2 = \frac{\psi''(\tau_0)}{2}(\tau - \tau_0)^2 \Rightarrow \mathrm{d}\tau = \sqrt{\frac{2}{\psi''(\tau_0)}}\mathrm{d}v, \tag{10}$$

we then update the variational (9) as

$$S(t, \hat{\omega}_i) = z(\tau_0)\sqrt{\frac{2}{\psi''(\tau_0)}}\int_{-\infty}^{\infty} \mathrm{e}^{jv^2}\mathrm{d}v. \tag{11}$$

According to the Fresnel integral $\int_{-\infty}^{\infty} \mathrm{e}^{jv^2}\mathrm{d}v = \sqrt{\frac{\pi}{2}} + j\sqrt{\frac{\pi}{2}}$, inserting the values of τ_0 and $\psi''(\tau_0)$, we obtain

$$S(t, \hat{\omega}_i) = a(t)\frac{1 + j}{2\sqrt{\pi\left[\hat{\omega}_i' + c(t)\right]\sigma_t}}. \tag{12}$$

The quality of using (12) as an approximation for CT is not only controlled by the item $c(t)$, but also by additional items depending upon more complex combinations of $\hat{\omega}_i$, σ_t and certain of their higher-order derivatives.

The main drawback of the TFR is that the different chirp rates $c(t)$ or Gaussian window width may not be simultaneously optimal for all the components. Considering a whale signal with the sampling frequency 48 kHz, the TFR of the signal is depicted in Fig. 4(a-f). If the components of UWA signals have different classes of TF structures (such as tone and pulse), it is necessary to use several space $(\triangle t, \triangle f)$ scales. A time-frequency-varying Gaussian kernel allows different window scales used at different time points.

B. TF-varying Gaussian window

The standard deviation of the underlying distribution can control the Gaussian window width, determined by [36]

$$L = 2\sigma\sqrt{2\ln 2}. \tag{13}$$

In [21], [36], an approximate relationship between the time-varying window width and the chirp rate has been derived. For a continuous signal, the local stationary length of $L(t)$ is defined by the chirp rate with the following condition:

$$L(t) = \max_{l} 2l \quad \text{s.t.} \int_{t-l}^{t+l} |\omega'|\mathrm{d}t \leq \Delta l, l > 0, \tag{14}$$

where $L(t)$ is adjusted by the threshold Δl such that the integral signal in $S(t, \omega)$ is quasi-stationary for each time instant t. The relationships in Eq.(14) cannot provide the optimal window width with a trade-off in parameter selection between the IF bias and σ. Thus, a more robust TF-varying standard deviation $\sigma(t, \omega)$ is designed to allow further control over the window width. According to Eq.(13) and Eq.(14), $\sigma(t, \omega)$ of Gaussian window is defined as

$$\sigma(t, \omega) = \frac{L(t, \omega)}{2\sqrt{2\ln 2}}, \tag{15}$$

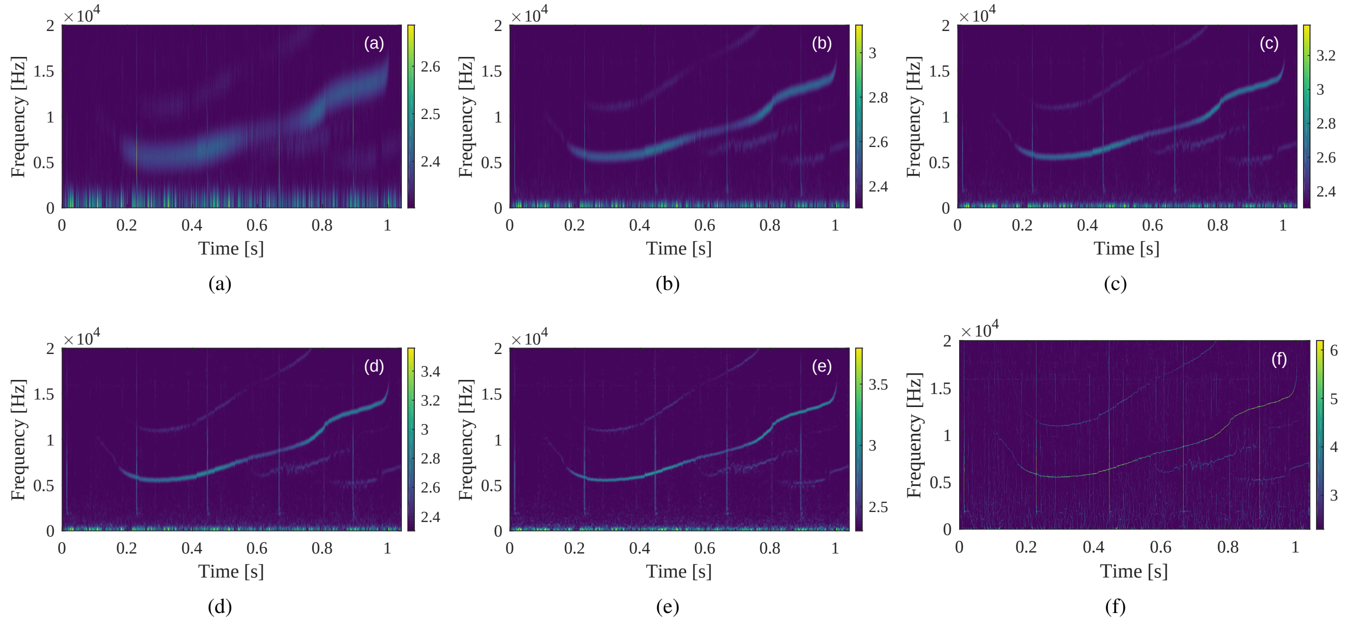


Fig. 4. TFR of an acoustic signal with different space scales. (a) $\triangle t = 15$, $\triangle\omega = 35$; (b) $\triangle t = 20$, $\triangle\omega = 30$; (c) $\triangle t = 25$, $\triangle\omega = 25$; (d) $\triangle t = 30$, $\triangle\omega = 20$; (e) $\triangle t = 35$, $\triangle\omega = 15$; (f) multiple space scales.

with

$$\text{s.t.} \begin{cases} \int_{t-l}^{t+l} |\hat{\omega}_i'(t,\omega)|\mathrm{d}t \leq \Delta l, & l > 0, \\ \int_{\omega-f}^{\omega+f} |W'(\omega_t,\omega_\omega)|\mathrm{d}\omega \leq \Delta f, & f > 0, \end{cases}$$

where $W'(\omega_t,\omega_\omega)$ is the Fourier transform of $\hat{\omega}_i'(t,\omega)$. The parameters Δl and Δf can be obtained respectively by the bounds of the frequency-modulated rates of the analytical signal.

Based on the TF-varying standard deviation, the Gaussian window [14] is generalized in TF space as

$$\hat{h}(t,\omega) = \frac{1}{\sqrt{2\pi}\sigma(t,\omega)} \mathrm{e}^{-\frac{\kappa}{2\sigma^2(t,\omega)}}, \tag{16}$$

with $\kappa = [t\ \omega]\boldsymbol{R}_{-\theta}[\lambda^2\ 0; 0\ \lambda^{-2}]\boldsymbol{R}_\theta[t\ \omega]^T$. $\boldsymbol{R}_\theta = [\cos\theta\ \ \sin\theta; -\sin\theta\ \ \cos\theta]$, where $\theta = \vartheta_{opt}$ is the optimal IRA of the chirp rate, $\lambda \geq 1$ indicates the anisotropic operator. When $\lambda = 1$, the TF-varying Gaussian window degenerates into a normal Gaussian window, and the optimal anisotropic operator is selected by maximizing SNR (Fig. 5). Here, we consider the whale signal whose TFR as shown in Fig. 4, where we add different noise, leading to different SNRs. Fig. 5 shows that for integer choices of λ. The highest level of concentration and Gaussianity is achieved as $\lambda = 2$. For a fixed height-to-width ratio, the chirplets have the highest level of TF energy in the range of $1.5 \leq \lambda \leq 2.5$.

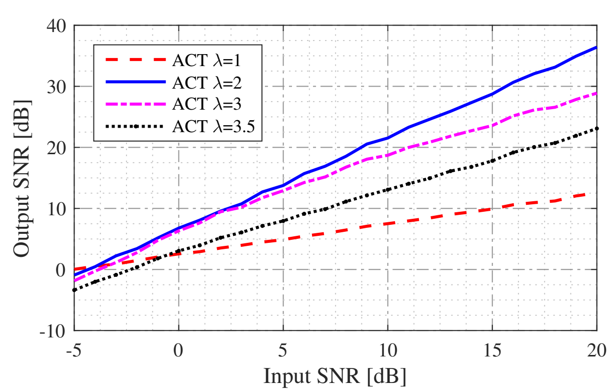


Fig. 5. Denoising results using different $\lambda = 1, 2, 3, 3.5$. The squared inner product between the Gaussian window approximation and chirplets with integer SNR from -5 to 20.

In the TF plane, the Gaussian window can be deemed as a two-dimensional Gaussian mask with the width $2\sqrt{2\ln 2}L_t$ and the height $2\sqrt{2\ln 2}L_\omega$. The height-to-width ratio μ is expressed as

$$\mu = \frac{2\sqrt{2\ln 2}L_t}{2\sqrt{2\ln 2}L_\omega} = \frac{\tan\theta}{2\pi\sigma^2}. \tag{17}$$

Accordingly, a global minimum of the envelope occurs [36] when $\partial\hat{h}(t,\omega)/\partial\sigma^2 = 0$, the TF-varying standard deviation is then obtained by $\sigma_{opt}^2 = 1/(2\pi|c(t)|)$. The TFR with the highest energy concentration can be implemented by the convolution of the ideal TFR with the 2D Gaussian mask $\mu = \tan\theta|c(t)|$.

When a noisy signal is corrupted by additive white Gaussian noise (AWGN) $\varepsilon(t,\omega)$ with a variance of σ_ε^2, it is filtered by TF-varying Gaussian window in (16). The noise suppression can be estimated by the noise variance in the filtered signal

$$\begin{aligned}\sigma_{\bar{\varepsilon}}^2 &= E\left\{\left[\varepsilon(t,\omega) * \hat{h}(t,\omega)\right]^2\right\}\\ &= \iint \hat{h}(\mathbf{v})\hat{h}(\bar{\mathbf{v}})E\left[\varepsilon(\mathbf{u}-\mathbf{v})\varepsilon(\mathbf{u}-\bar{\mathbf{v}})\right]\mathrm{d}\mathbf{v}\mathrm{d}\bar{\mathbf{v}}\\ &= \sigma_\varepsilon^2 \iint \hat{h}(\mathbf{v})\hat{h}(\bar{\mathbf{v}})\delta(\mathbf{v}-\bar{\mathbf{v}})\mathrm{d}\mathbf{v}\mathrm{d}\bar{\mathbf{v}}\\ &= \sigma_\varepsilon^2 \iint \left[\hat{h}(\mathbf{v})\right]^2\mathrm{d}\mathbf{v} = \frac{\sigma_\varepsilon^2}{2\pi\sigma^2},\end{aligned} \tag{18}$$

where $\mathbf{u} = [t,\omega]^T$, $\sigma_{\bar{\varepsilon}}^2$ is the variance of a smoothed noise. (18) demonstrates that the capability of noise suppression is independent of the anisotropic operator and preference orientation.

In this case, the ACT only compensates the energy of available signals, but the energy of noise is not concentrated. Thus, the proposed method is well suitable to IF estimation in an environment of high level noise.

III. Multimodal sparse time-frequency representation

A. Directional Chirplet Ridges

Generally, the directional chirplet ridges with directionality points $(\hat{t}_i(t,\omega), \hat{\omega}_i(t,\omega))$ satisfy $\hat{\omega}_i(t,\omega) = \omega$ or equivalently

$$\Im\left\{\frac{\varrho}{S}e^{j\vartheta(t)}\right\} = 0. \tag{19}$$

If defining directional ridges by linking together the points that satisfy (19), splitting IF at the zeros of ACT will guarantee smooth continuity of the phase derivative along every ridge segments.

For precise parameter estimation, a two-dimensional candidate IF $\hat{\omega}_i(t,\omega)$ [13] for the ACT can be expressed as

$$\hat{\omega}_i(t,\omega) = -j\frac{\partial_t S(t,\omega)}{S(t,\omega)}. \tag{20}$$

The partial derivative can be calculated as

$$\begin{aligned}\partial_t S(t,\omega) &= \partial_t\left(\int_{-\infty}^{\infty} s(\tau)\hat{h}(\tau-t,\omega)\mathrm{e}^{j\frac{c}{2}(\tau-t)^2}\mathrm{e}^{-j2\pi\omega\tau}\mathrm{d}\tau\right)\\ &= -S_{\hat{h}'}(t,\omega) + jc(t-\tau)S(t,\omega),\end{aligned} \tag{21}$$

where $S_{\hat{h}'}(t,\omega) = S(t,\omega)\partial_t\hat{h}(t,\omega)$, and $\partial_t\hat{h}(t,\omega)$ is a first partial derivative of TF-varying Gaussian window in the direction θ, obtained by

$$\partial_t\hat{h}(t,\omega) = \partial_t\hat{h}(\boldsymbol{R}_\theta\boldsymbol{u}) = \frac{[\cos\theta,\sin\theta]\boldsymbol{u}}{\lambda^{-2}\sigma^2}\hat{h}(t,\omega). \tag{22}$$

According to (20)-(22), the IF for ACT is then rewritten as

$$\hat{\omega}_i(t,\omega) = j\frac{(t\cos\theta + \omega\sin\theta)}{\lambda^{-2}\sigma^2}\hat{h}(t,\omega) + jc(t-\tau). \tag{23}$$

This idea follows the curvilinear structures definition by Steger [37], and identifies ridges from the analysis of the Hessian of the multimodal signals. Thus, the second derivative of a phase in the direction θ is expressed as

$$\hat{\omega}_i'(t,\omega) = -j\frac{\lambda^2}{\sigma^2}\left(\frac{(t\cos\theta+\omega\sin\theta)^2}{\lambda^{-2}\sigma^2} - 1\right)\hat{h}(t,\omega) + jc. \tag{24}$$

The $\hat{\omega}_i'(t,\omega)$ considers a variety of ridge curvatures for adapting to the local modes of the TFR. The resulting ridge representation of the signal in direction angle θ is given by points (t,ω) satisfying

$$\omega - \omega_0 = \hat{\omega}_i'(t,\omega)(t-t_0). \tag{25}$$

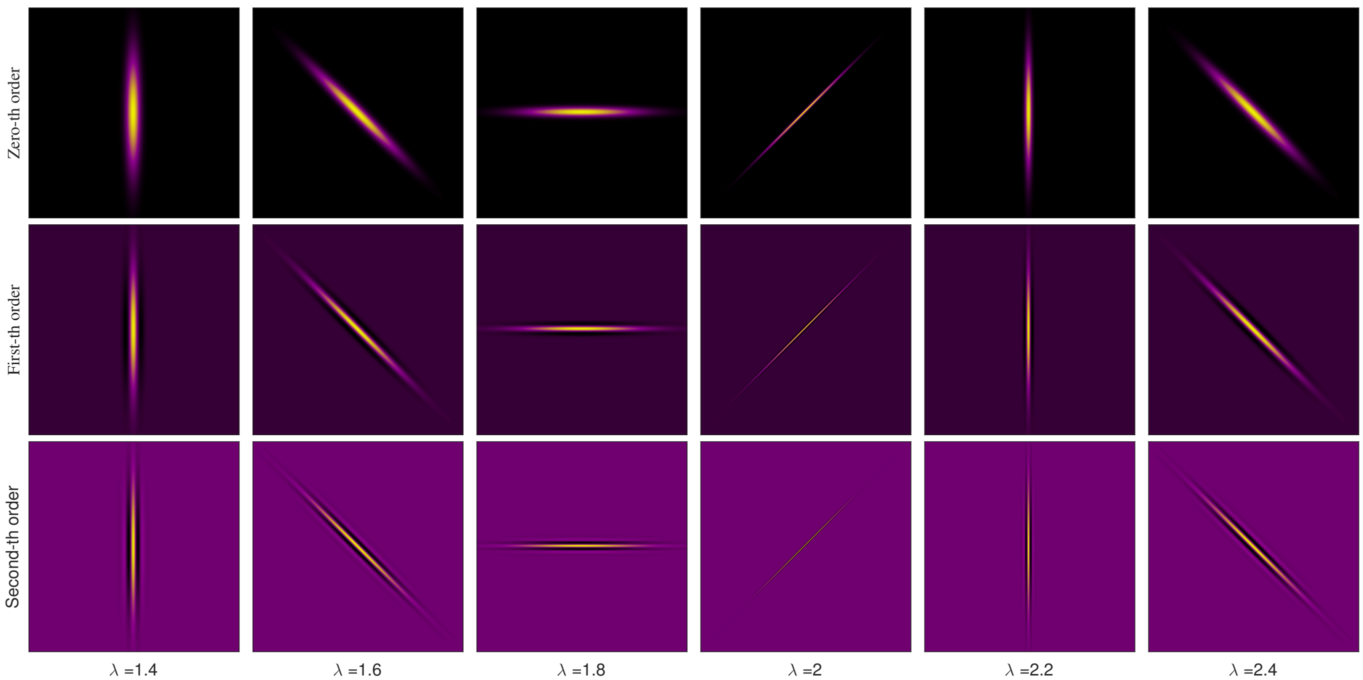


Fig. 6. TF representation of different orders $\hat{\omega}_i(t, \omega)$ featuring various parameter configurations. For each value of λ, the windows with $\theta = \frac{\pi}{2}, \frac{3\pi}{2}, 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{2}$ (left to right) are displayed, respectively.

The IF can be estimated unbiasedly by the chirplet ridges while only need to satisfy $\tan\theta = \hat{\omega}_i'(t, \omega)$.

Fig. 6 contains several examples of TFR with different orders $\hat{\omega}_i(t, \omega)$. Since the approach uses the second directional derivatives of TFR for the extraction of the line points in the ridges, no specialized directional filters are needed. If the ridges detection is not perfect, (25) can overcome the discontinuity caused by noise in the ridges. TF-varying Gaussian window provides a simpler and more powerful approach for the study of multimodal structures and empowering the representativity of the properties of the patterns (orientation, curvature, scale). Thus, the algorithm can extract ridges of different curvatures.

B. Structure-split-merge algorithm

As ideal acoustic components generally present sparsity in a certain TF domain, structure-split-merge method, together with the aid of TF methods, can effectively recover the tonal and pulsed features of the signal. The fixed choice of $(\hat{t}_i, \hat{\omega}_i)$ from the TFR determines a split-merge transformation kernel, satisfying detailed balance. Selecting a main $(\hat{t}_i, \hat{\omega}_i)$ pair that interprets similar data is important for efficiency.

The output of ACT is a TFR that contains real parts and imaginary parts. SSM, which splits the time-frequency space into contours of interest, utilizes this property by counting the number Σ_S of detections set to energy value in sub-contours of the energy spectrogram. Therefore, the reconstruction procedure for multimodal sparse representation based on SSM is introduced in Algorithm 1.

The hypotheses $\kappa_v = \Sigma_v$ contains imaginary parts only and $\kappa_{rv} = \Sigma_r + \Sigma_v$ contains real and imaginary parts. The observed energy function ρ_{κ_v} and $\rho_{\kappa_{rv}}$ can be obtained with the entire spectrogram. Given a sparse distribution, an estimate of detections is determined by $\rho_\Sigma = -\{\rho_{\kappa_v} < 0\} + \{\rho_{\kappa_{rv}} > 0\}$. As a split criterion, the estimate $\rho_{\kappa_{rv}}$ is provided by minimization of the local KullbackLeibler divergence [38] between the observed distribution ρ_{Σ_S} and the TF distribution $S(t, \omega)$ obtained by ACT. The merge criterion is constructed based on the two distribution. Therefore, one can determine the merge sparse TF representation $\hat{S} = \Sigma\rho_{\Sigma_S} \times |S(t, \omega)|$.

The sound example, consisting of a mix of whistle analysis, is shown in Fig. 7. In the case of a whale whistle, the TFR (Fig. 7(a)) information prepares an analytical characterization of signal features. It is easy to notice from the observation of the TFR that noise exists in the spectrogram representation. Observing that the SSM can eliminate the noise from the TF spectrogram, the noise robustness of the SSM method is illustrated in Fig. 7(b). Unlike the spectrogram, the estimated information is clearly presented because only the estimated valuable components are plotted.

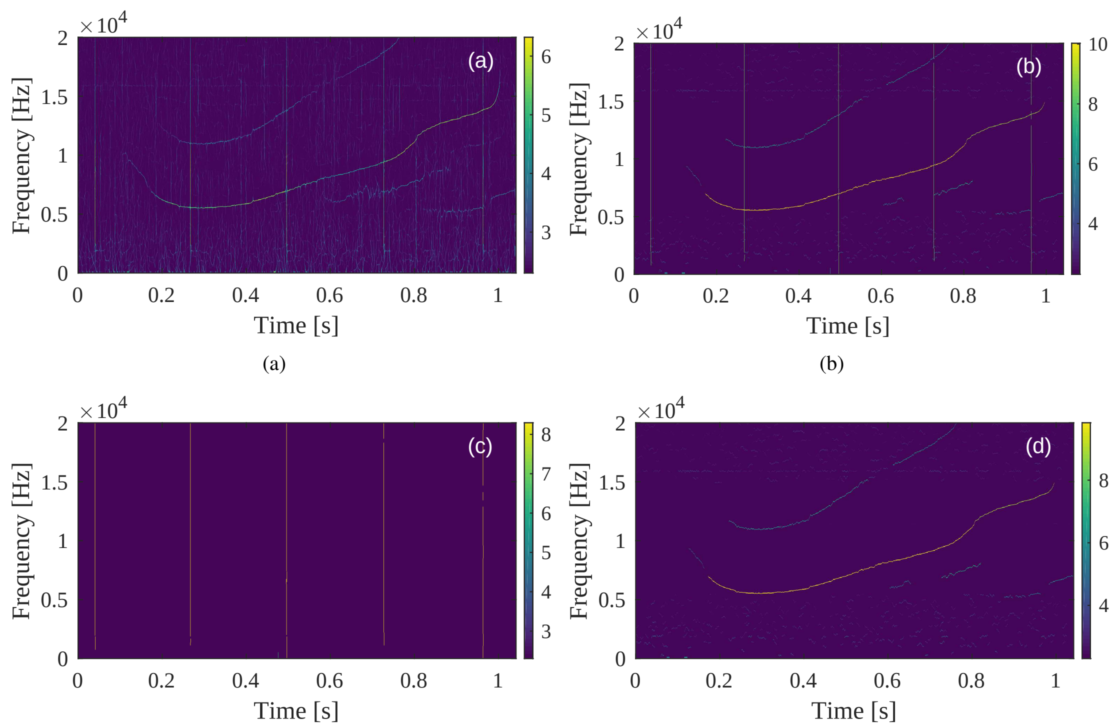
The information of the SSM contains the modes of the pulsed transients (Fig. 7(c)) and the tones (Fig. 7(d)). These split modes could have an amount of applications, such as signal detection and classification, underwater source separation and reconstruction, UWA communication signal characterization, etc.

Algorithm 1: SSM algorithm

Input: Time-frequency representation S, the second derivative $\hat{\omega}'(t,\omega)$ of a signal phase in the direction ϑ;

Output: Multimodal sparse representation $\hat{S}$;

Initialize the spectrum S, ϱ;
Initialize $\eta = \hat{\omega}' \cdot S$;
$N = \text{length}(\vartheta)$;
for $k = 1 : N$ **do**
  $\rho_{\Sigma} = -\{(1-\rho_{\kappa_{rv}}) < 0\} + \{\rho_{\kappa_{rv}} > 0\}$;
  $g_t = \partial_t S,\ g_\omega = \partial_\omega S$;
  $\Delta S = -g_t \cos(\vartheta(k) + \frac{\pi}{2}) + g_\omega \sin(\vartheta(k) + \frac{\pi}{2})$;
  $\mu = \text{Split}(\Delta S, \eta, \varrho)$;
  $l = \max(\mu, 2)$;
  $w = []$,$M = \text{length}(l)$;
  for $i = 1 : M$ **do**
    $w(i) = l(i)$
  end
  $\hat{S} = \text{Merge}(w, S)$
end
return $\hat{S}$;



(a) (b) (c) (d)

Fig. 7. TFR of a whale signal with several harmonics, and pulsed bursts. (a) TFR; (b) multimodal sparse representation; (c) extracted pulsed components; (d) extracted tonal components.

IV. Application to UWA signals

To verify the effectiveness of the proposed technique, we apply field experimental UWA communication signals and whale signals shown as below.

***Example* 1:** To analyze the performance of the proposed ACT on the energy concentration, the ACT is compared with the CT. For the ACT, the initial angle list ranges from $\frac{\pi}{4}$ to π, and an anisotropic operator is $\lambda = 2$. A Hanning window with length 512 is employed in the spectrogram computation of the CT. We evaluate the ACT with an UWA communication dataset.

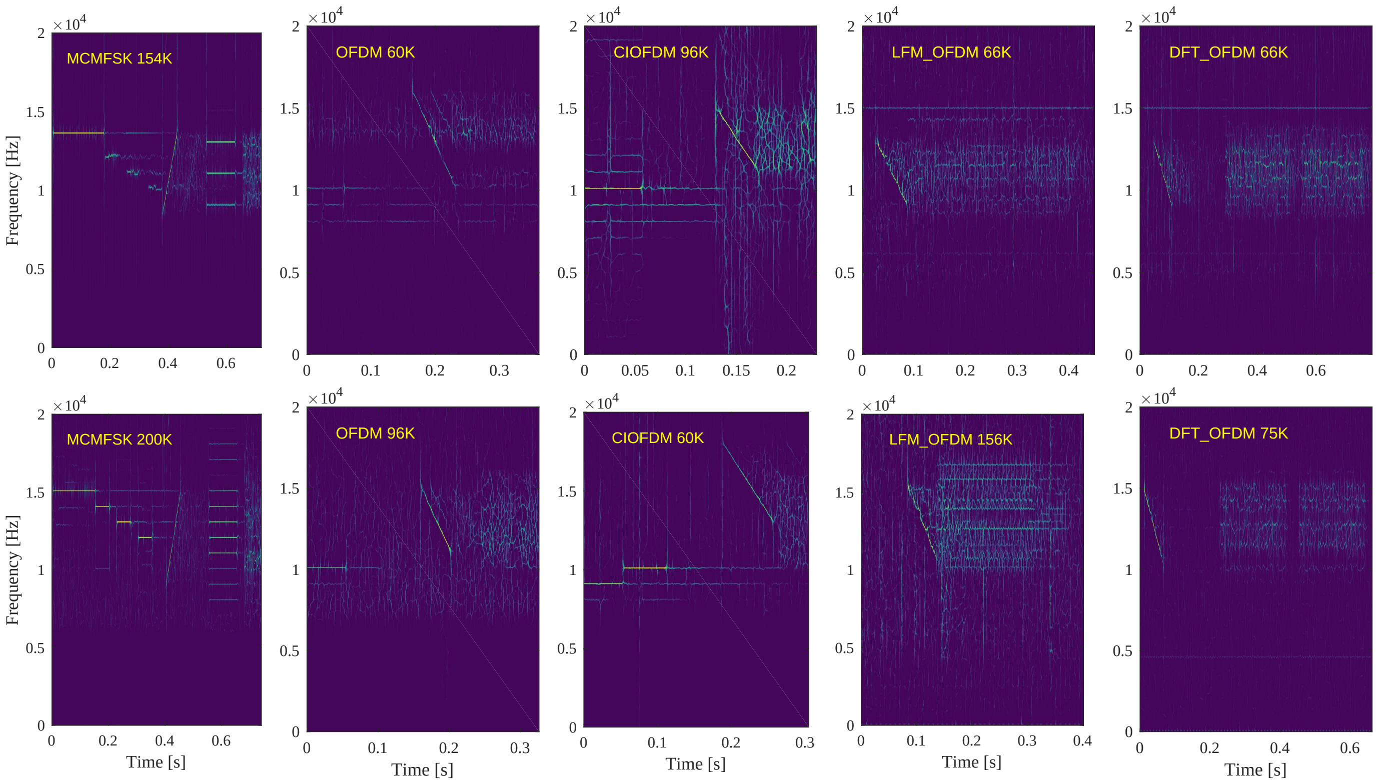


Fig. 8. TFRs of UWA communication signals by the ACT.

Different datasets with different modulations were collected at different locations in Wuyuan Bay, Xiamen, China, from 2016 to 2018. We use sensors called *Universal Deck Device* and *Underwater Acoustic Transducer*, the specifications of which are listed in Table. I. Dataset consists of audio files (.WAV) with length between 2.6 s and 42.3 s.

TABLE I
Specifications of two sensors.

Description	Parameters
Universal Deck Device (UDD-630 Series)	• Usable frequency: 20 to 40 kHz • Highest bit rate: 2.4 kbps • Maximum cable length: 200 m • Interface type: RS-232 • Directivity: Omni
Underwater Acoustic Transducer	• Resonant frequency: 30 kHz • Usable frequency: 20 to 40 kHz • Maximum operating depth: 300 m • Horizontal: Omni-directional • Vertical: 280°

Fig. 8 displays the spectrogram of the five modulation types. They are labelled as MCMFSK, OFDM (Orthogonal Frequency Division Multiplexing), CIOFDM (Carrier Interferometry, CI), LFM_OFDM (Linear Frequency Modulation, LFM) and DFT_OFDM (Discrete Fourier Transform, DFT), respectively. Each signal contains guide fragment and data fragment. For the variety of the recorded communication signals, the guide fragments present very different structures due to the superimposition

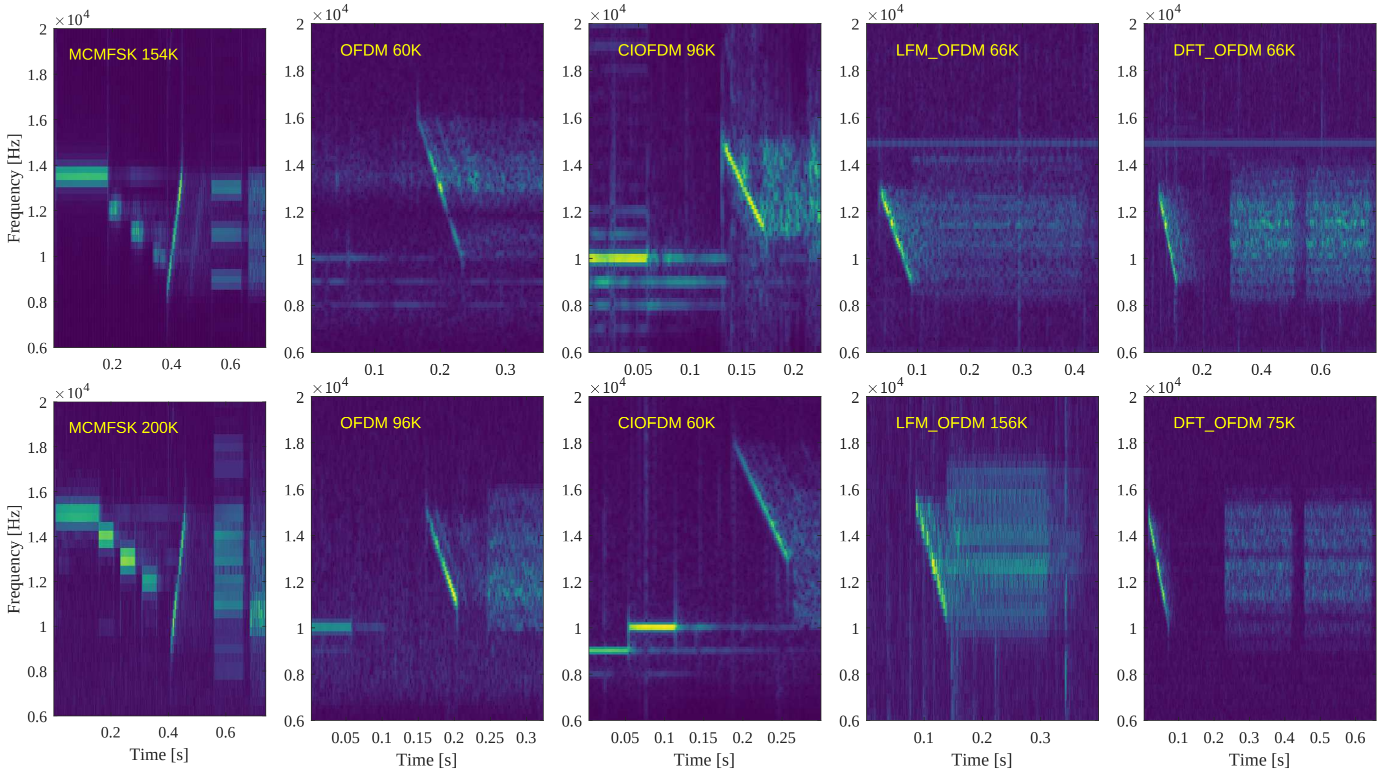


Fig. 9. TFRs of UWA communication signals by the CT.

of linear frequency modulation components. Since these structures have different chirp rates, it is better to use time-frequency-varying standard deviation adapted to each component.

In Fig. 8, the ten spectrogram plots describe the variations of time-frequency contour in terms of their shape, energy distribution, time duration and frequency span. Some components may be blurred, as shown in Fig. 9, but a few representative TF structures of the signal are extracted by the CT. The CT suffers from poor energy concentration for two main reasons: 1) the window is time-varying but not frequency-varying; 2) the relationship between the window width and the chirp rate in (17) is not adequate. This example verifies that the ACT is superior to the CT for UWA communication signals.

The information obtained by the proposed method is rich, because we obtain a series of parameters, such as modulation type, bandwidth, duration, location in time, etc. The proposed method is a valuable tool in the analyzed signals involved in energy concentrated in the components. In this case, a separation of the guide fragment makes the communication type easily interpretable. Thus, through the TF representation of the modulation signals, it is easy to be discriminated based on the two parts.

***Example* 2:** Compared to the above example, we show a more complicated example here, where the analysed signal consists of multiple components with pulsed and tonal components overlapping in the TF plane. The carried out example shows that, apart from the sparse TF representation, SSM-based PTR analysis determines if the signal has a dominant pulsed or tonal component.

In terms of whale signals, the detection of multiple events depends on pulsed and tonal components as well as a combination of these two main categories. However, there are some situations, where the algorithm using only TFR features, do not guarantee a correct detection of acoustic events. The TFR with the aid of the pulsed-to-tonal ratio (PTR) [3] to distinguish between two similar whale signals. When a 'click' event appear in a signal, the value of PTR increases, but the PTR decreases as the number of tone events increases. Therefore, the PTR provides valuable information for comparing signals that contain energy which is concentrated in the tonal or pulsed components. In the paper, an alternative ratio can also be derived by decomposing the TFR using the SSM. The PTR can be calculated as

$$R = 10 \log \frac{S(t,\omega)_p}{S(t,\omega)_t}, \tag{26}$$

where $S(t,\omega)_p = \sum_{m=1}^{M} S(t,m)_p$ and $S(t,\omega)_t = \sum_{m=1}^{M} S(t,m)_t$ are sum of the spectrum of pulsed and tonal components at the interval t, respectively.

In the test, we consider two marine-mammal signals that were recorded at the *Voices in the sea* website. The sample frequency of both the selected signals of length N_s =20000 is equal to 48 kHz. Some chirp-like components correspond to the associated non-linear tonal components existing in the signal structure of a ringtone Bottlenose.

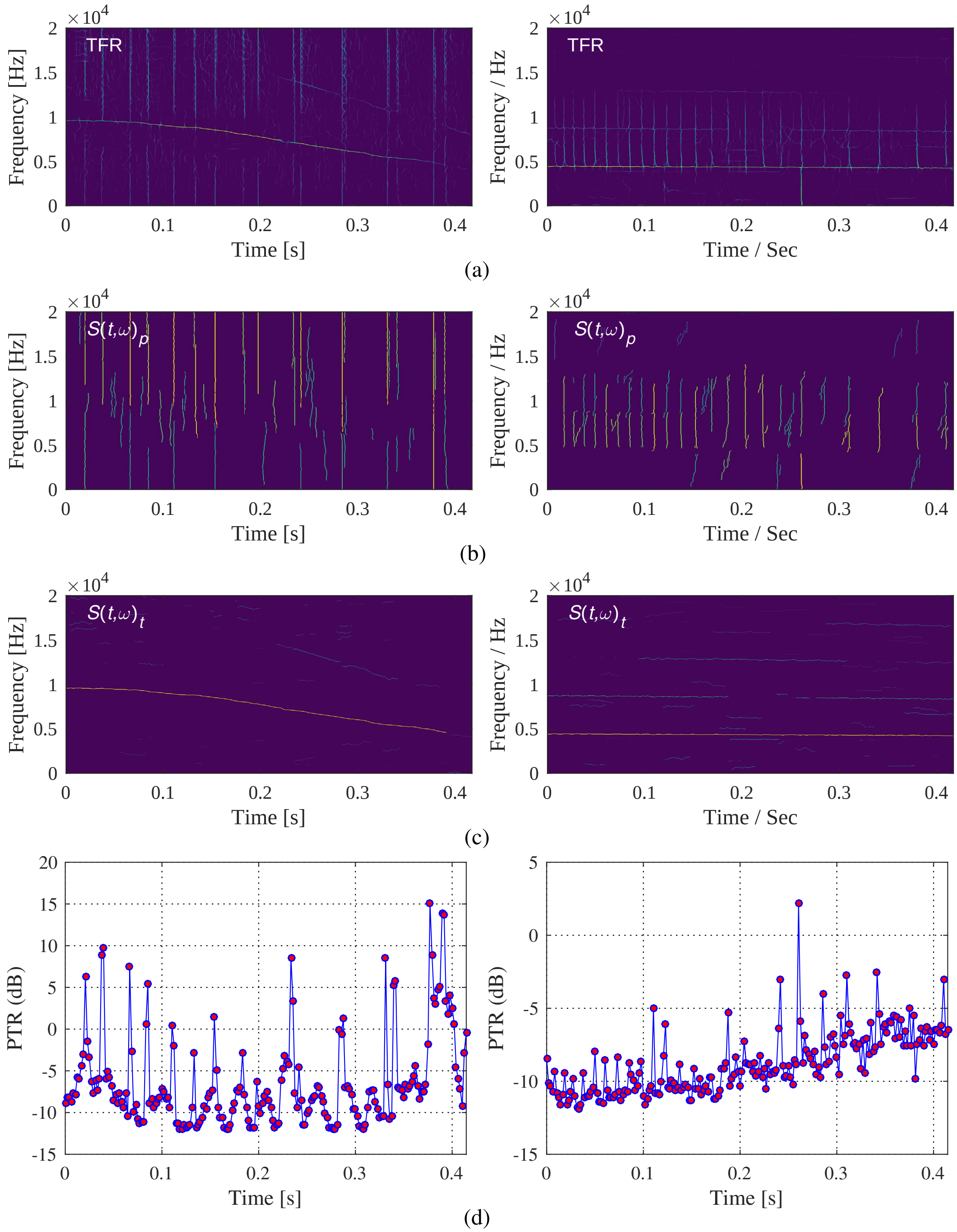


Fig. 10. TFR and PTR estimation of two whale signals, ringtone Bottlenose (left) and ringtone Beluga (right). (a) TFR by ACT. (b) Pulsed components by SSM. (c) Tonal components by SSM. (d) Estimated PTR.

In Fig. 10, we observe that the TF components represented by the ACT algorithm are visually close to the acoustic event behaviour illustrated by the TF spectrogram. By unifying the arbitrary curvatures for all frequency sub-bands and for all temporal location, the detection ridges are obtained by the SSM in both time and frequency domains. This IT and IF ridges cover information of the transient pulsed and tonal components of the signals. The analysis of extracted chirplets and corresponding modes from all contours of interest leads to the identification of TF components of the signal. Therefore, the sparse modeling of signal can provide the parametric information about the signal. The extracted parameters can be related to the signal features. After the sparse representation, for each corresponding signals, the feature extraction procedure is executed to obtain the parameters PTR of TF components.

Fig. 10 also displays a comparison of the PTR when multimodal components are decomposed into IF $S(t,\omega)_t$ and IT

$S(t,\omega)_p$ modes of the tonal and pulsed components using the SSM algorithm. The proposed SSM-based estimation provides a higher sensitivity to small transient pulses in low PTR signals. The result illustrates that the PTR increases as the density of bursts increases.

Moreover, the analytical estimation of the PTR provides useful information for further processing stage. The parameters of the PTR constitute the extracted patterns of the analysed signal and they could be used for signal detection or classification.

***Example* 3:** Besides the energy concentration, we measure the time complexity of different TFR methods on the Intel(R) Core(TM) i7-8700 CPU and MATLAB R2019b. Table. II compares the computational complexity of the proposed method to the CT and Fourier synchrosqueezing transform (FSST) [25] on two datasets in the Example 1 and 2.

Compared to the CT, the FSST has much higher complexity due to its reassignment process to improve the readability of the TFR. The proposed method has lower complexity than FSST, because the calculation of the SSM in the ACT is simpler than the optimization process in the FSST. Further, the proposed approach is more flexible, allowing an adjustment between the complexity and the energy concentration.

TABLE II
COMPARISON COMPLEXITY OF DIFFERENT METHODS.

Methods	MCMFSK	OFDM	CIOFDM	LFM_OFDM	DFT_OFDM	Whale
CT	0.17	0.038	0.032	0.047	0.12	0.051
FSST [25]	5.3	2.58	1.81	2.02	3.08	2.42
ACT	2.86	1.06	1.03	1.19	2.05	1.13
ACT with SSM	2.98	1.51	1.39	1.5	2.37	1.56

V. CONCLUSIONS AND DISCUSSION

In this paper, we propose the ACT to generate high-resolution TF representations for the analysis of UWA signals. The ACT with the TF-varying Gaussian window allows the use of different window width controlled by anisotropic operators at different time points. Its advantage lies in its ability to achieve high energy concentration and readable TF representation without auto-term distortion, demonstrated by the application of UWA communication signals. Better time-frequency resolution than the CT can be attained by the ACT, although at a higher computational cost. A comparison with the reassigned FSST shows that the ACT gives better results with less computational complexity. The ACT, as a kind of general tools for the analysis of overlapping multicomponent signals, will bring more advanced applications in a future research.

For a classification, a critical step is to extract discriminating features from the TFR. The directional chirplet ridges from the TFR can provides the estimation of intersected IF and IT of multicomponent signals. Utilizing the chirplet ridges, the proposed SSM can decompose and extract the tonal and pulsed features while is regarded as a filter to suppress noise. The multimodal sparse representation obtained by the SSM with similar properties to the ACT, the improvement in performance over the ACT has been verified through the analysis of whales. A new parameter PTR can measure the pulsed to tonal strength of multimodal sparse representation. The SSM-based PTR gives information about the energy in the pulsed component in relation to the energy in the tonal component of a given signal. A real example has illustrated the utility of the parameter in helping to classify whale sounds with mixed pulses and tonal modes.

Moreover, the presented methodology may potentially be useful in suppressing reverberation. Consequently, the proposed sparse representation methodology, as a standard signal imaging techniques, could potentially be merged in the learning model to further enhance the concurrent detection, identification and localization of underwater multi-target. The application of SSM on UWA signals can exact time-frequency features to monitor the structural state of acoustic events, which effectively achieve online extraction of the structural signature.

ACKNOWLEDGMENT

The work of Yongchun Miao and Haixin Sun was partly supported by the National Key R&D Program of China(2018YFC080 9200) and the National Natural Science Foundation of China (61671394) and the Fundamental Research Funds for the Central Universities (20720170044). The work of Jianghui Li was partly supported by the European Union's Horizon 2020 research and innovation programme under the grant agreement number 654462 (STEMM-CCS).

REFERENCES

[1] C. Ioana, A. Quinquis, and Y. Stephan, "Feature extraction from underwater signals using time-frequency warping operators," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 3, pp. 628–645, July 2006.

[2] Y. Y. Al-Aboosi and A. Z. Sha'ameri, "Improved underwater signal detection using efficient timefrequency de-noising technique and pre-whitening filter," *Applied Acoustics*, vol. 123, pp. 93–106, August 2017.

[3] R. Miralles, G. Lara, J. A. Esteban, and A. Rodriguez, "The pulsed to tonal strength parameter and its importance in characterizing and classifying beluga whale sounds," *Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2173–2179, March 2012.

[4] J. Li, P. R. White, J. M. Bull, and T. G. Leighton, "A noise impact assessment model for passive acoustic measurements of seabed gas fluxes," *Ocean Engineering*, vol. 183, no. 1, pp. 294–304, 2019.

[5] J. Li and Y. V. Zakharov, "Efficient use of space-time clustering for underwater acoustic communications," *IEEE Journal of Oceanic Engineering*, vol. 43, no. 1, pp. 173–183, 2018.
[6] J. Li, Y. V. Zakharov, and B. Henson, "Multibranch Autocorrelation Method for Doppler Estimation in Underwater Acoustic Channels," *IEEE Journal of Oceanic Engineering*, vol. 43, no. 4, pp. 1099 – 1113, 2018.
[7] F. Yuan, Z. Jia, J. Li, and E. Cheng, "STLFM Signal Based Adaptive Synchronization for Underwater Acoustic Communications," *IEEE Access*, vol. 7, pp. 28 734–28 748, 2019.
[8] Y. Zhang, T. Wu, Y. Zakharov, and J. Li, "MMP-DCD-CV based sparse channel estimation algorithm for underwater acoustic transform domain communication system," *Applied Acoustics*, vol. 154, pp. 43–52, 2019.
[9] B. G. Mobasseri and R. S. Lynch, "Information embedding in sonar by modifications of time-frequency properties," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 1, pp. 139–154, Jan. 2016.
[10] B. G. Ferguson, "Time-frequency signal analysis of hydrophone data," *IEEE Journal of Oceanic Engineering*, vol. 21, no. 4, pp. 537–544, 1996.
[11] J. Li, P. R. White, and B. Roche, "Seafloor noise ensemble from vessel manoeuvre in the central North Sea," *Ocean Engineering*, vol. 196, p. 106836, 2020.
[12] J. Li, P. R. White, B. Roche, J. W. Davis, and T. G. Leighton, "Underwater radiated noise from hydrofoils in coastal water," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3552–3561, 2019.
[13] N. A. Khan and B. Boashash, "Instantaneous frequency estimation of multicomponent nonstationary signals using multiview time-frequency distributions based on the adaptive fractional spectrogram," *IEEE SIGNAL PROCESSING LETTERS*, vol. 20, no. 2, pp. 157–160, Feb 2013.
[14] T. Oberlin, S. Meignen, and V. Perrier, "Second-order synchrosqueezing transform or invertible reassignment? towards ideal time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1335–1344, March 2015.
[15] H. K. Kwok and D. L. Jones, "Improved instantaneous frequency estimation using an adaptive short-time fourier transform," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2964–2972, October 2000.
[16] J. M. Lilly and S. C. Olhede, "On the analytic wavelet transform," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 4135–4156, August 2010.
[17] H. Lee and S. Kwon, "Wave profile measurement by wavelet transform," *Ocean Engineering*, vol. 30, no. 18, pp. 2313–2328, 2003.
[18] W. Amin, M. Davis, G. Thomas, and D. Holloway, "Analysis of wave slam induced hull vibrations using continuous wavelet transforms," *Ocean Engineering*, vol. 58, pp. 154–166, 2013.
[19] B. Boashash and P. O'Shea, "Use of the cross wigner-ville distribution for estimation of instantaneous frequency," *IEEE Transactions on Signal Processing*, vol. 41, no. 3, pp. 1439–1445, March 1993.
[20] R. Imai, Y. Hashimoto, K. Kikuchi, and S. Fujii, "High-resolution beamforming by the Wigner-Ville distribution method," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 1, pp. 105–110, 2000.
[21] S. Mann and S. Haykin, "Adaptive "chirplet" transform: an adaptive generalization of the wavelet transform," *Optcal Engineering*, vol. 31, no. 6, pp. 1243–1256, June 1992.
[22] ——, "The chirplet transform: physical considerations," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2745–2761, November 1995.
[23] Y. Miao, H. Sun, and J. Qi, "Synchro-compensating chirplet transform," *IEEE Signal Processing Letters*, vol. 25, no. 9, pp. 1413–1417, September 2018.
[24] H. Zou, Y. Chen, J. Zhu, Q. Dai, G. Wu, and Y. Li, "Steady-motion-based Dopplerlet transform: Application to the estimation of range and speed of a moving sound source," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 3, pp. 887–905, 2004.
[25] A. Ahrabian and D. P. Mandic, "A class of multivariate denoising algorithms based on synchrosqueezing," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2196–2208, May 2015.
[26] F. Auger, P. Flandrin, Y. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H. Wu, "Time-frequency reassignment and synchrosqueezing: An overview," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 32–41, November 2013.
[27] L. Stankovi, "A measure of some timefrequency distributions concentration," *Signal Processing*, vol. 81, no. 3, pp. 621 – 631, March 2001.
[28] S. Chen, X. Dong, Y. Yang, W. Zhang, Z. Peng, and G. Meng, "Chirplet path fusion for the analysis of time-varying frequency-modulated signals," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1370–1380, February 2017.
[29] I. Djurović, M. Simeunović, and P. Wang, "Cubic phase function: A simple solution to polynomial phase signal analysis," *Signal Processing*, vol. 135, pp. 48–66, June 2017.
[30] D. A. Cook and D. C. Brown, "Analysis of phase error effects on stripmap SAS," *IEEE Journal of Oceanic Engineering*, vol. 34, no. 3, pp. 250–261, 2008.
[31] K. Pelekanakis and M. Chitre, "Robust equalization of mobile underwater acoustic channels," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 4, pp. 775–784, 2015.
[32] T. Yang, C.-F. Huang, S. Huang, and J.-Y. Liu, "Frequency striations induced by moving nonlinear internal waves and applications," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 3, pp. 663–671, 2016.
[33] M. Aoi, K. Lepage, Y. Lim, U. T. Eden, and T. J. Gardner, "An approach to time-frequency analysis with ridges of the continuous chirplet transform," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 699–710, February 2015.
[34] G. Yu and Y. Zhou, "General linear chirplet transform," *Mechanical Systems and Signal Processing*, vol. 70-71, pp. 958 – 973, March 2016.
[35] B. Mulgrew, "The stationary phase approximation, time-frequency decomposition and auditory processing," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 56–68, January 2014.
[36] R. A. Carmona, W. L. Hwang, and B. Torresani, "Multiridge detection and time-frequency reconstruction," *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 480–492, February 1999.
[37] C. Steger, "An unbiased detector of curvilinear structures," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 20, no. 2, pp. 113–125, February 1998.
[38] T. van Erven and P. Harremos, "Rnyi divergence and kullback-leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.