Signal Processing 90 (2010) 1591–1596

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

High resolution spectrograms using a component optimized short-term fractional Fourier transform

Aled T. Catherall *, Duncan P. Williams

Defence Science & Technology Laboratory, Physical Sciences, Porton Down, Salisbury SP4 0JQ, UK

ARTICLE INFO

Article history:
Received 18 March 2009
Received in revised form
22 October 2009
Accepted 3 November 2009
Available online 11 November 2009

Keywords:
Fractional
Fourier
Spectrogram
Time–frequency

ABSTRACT

We present an algorithm based on a short-term representation of the fractional Fourier transform which is highly suited to signals that contain multiple non-stationary components. To compare the time–frequency resolution of such signals using our proposed technique, we analyze several signals which contain multiple non-stationary components, including a synthetic signal and a bat echolocation signal. We demonstrate that our technique offers time–frequency representations with superior time–frequency resolution than several other established methods.

1. Introduction

The invention of the spectrogram in the 1940s for the analysis of speech has gone on to find many other applications. The thousands of articles and books that apply the spectrogram to measure the time and frequency information of a signal are all evidence to how widespread it has been and can be utilized beyond speech sounds. For example, non-stationary signals are analyzed in radar and sonar applications [1,2] to track targets or in medical ultrasound to measure blood flow [3]. In these and many other examples it is important to assemble a high resolution spectrogram so that the true time and frequency characteristics of the signal can be measured and, for example, an appropriate decision or course of action taken that is based on the information.

The spectrogram is typically calculated from the intensity of the short-term Fourier transform (STFT) [4] that is the running output of the Fourier transform of a signal. The traditional means to calculate the STFT (and spectrogram) is to window a signal (e.g. time series) into discrete intervals, which are each Fourier transformed in turn to obtain the local power spectrum of the signal. The problem with measuring the spectrogram this way is that it is not always well suited to the analysis of non-stationary signals. It is well known that long analysis windows result in a poor time resolution whilst short analysis windows cannot always achieve the desired frequency resolution.

It is for these reasons that other ways to measure the spectrogram have been widely studied [4,5]. Methods such as wavelets [6], Wigner distributions and its modifications [7,8], and reassignment methods (e.g. [9]) have found some success for different applications. However, these other methods have not always been widely adopted by the signal processing community since they may be difficult to interpret, may suffer from the appearance of unwanted cross-terms, or may be cluttered with meaningless random points. This can limit the resolution of the spectrogram or disguise the time and frequency information of a signal with other misleading information.

In recent years, there has been a considerable amount of attention on the fractional Fourier transform (FrFT) by

* Corresponding author.
E-mail address: atcatherall@dstl.gov.uk (A.T. Catherall).

0165-1684/$ - see front matter
doi:10.1016/j.sigpro.2009.11.004

the signal processing community. One of the most interesting applications is the possibility of higher resolution analysis of non-stationary signals [10–12]. Methods such as those described by [10] have not been found to be suited to signals containing multiple non-stationary components and, in particular, signals containing multiple non-related chirps.

This paper is devoted to a new method based on the fractional Fourier transform which is ideally suited to signals which contain several non-stationary (or moving) components. In this method each individual non-stationary component in a signal is identified, and then each one is processed independently and in turn to produce a spectrogram with a high time–frequency resolution. To compare our method against other existing methods for the time–frequency representation of such signals, we analyze an artificial multi-component signal and a real bat echolocation signal. The results obtained from each processing scheme are displayed and discussed.

2. Transforms

2.1. The fractional Fourier transform (FrFT)

The Fourier transform can be interpreted as a rotation by $\pi/2$ in the time–frequency plane, since two successive Fourier transform operations on a signal yield a time-reversed replica of the original signal [13]. The fractional Fourier transform is a generalization of the Fourier transform, and may be regarded as a rotation in the time–frequency plane by any arbitrary angle ϕ. The FrFT of a function $f(t)$ is given by

$$F_\alpha(x) = \frac{e^{-\mathrm{i}(1/4\pi\,\mathrm{sgn}(\phi)-1/2\phi)}}{(2\pi|\sin\phi|)^{1/2}} \exp\left[\frac{1}{2}\mathrm{i}x^2\cot\phi\right] \times \int_{-\infty}^{\infty} \exp\left[-\mathrm{i}\frac{xt}{\sin\phi} + \frac{1}{2}\mathrm{i}t^2\cot\phi\right] f(t)\,\mathrm{d}t, \quad (1)$$

where $\alpha = (2/\pi)\phi$ is the transform order, t is time, $\mathrm{i} = \sqrt{-1}$ and x is the transform domain. When $\alpha = 1$, the transform reduces to the standard Fourier transform and x represents frequency. When $\alpha = 0$ the transform reduces to the identity operator. The FrFT of a linear chirp of the form $e^{\mathrm{i}(at^2+bt+c)}$ yields a delta function when the value of α is selected to match the chirp rate

$$\alpha = -\frac{2}{\pi}\tan^{-1}\left(\frac{1}{2a}\right). \quad (2)$$

The position of the peak at the optimal fractional order is related to its center frequency b

$$x = b\sin\left(\alpha\frac{\pi}{2}\right). \quad (3)$$

A variety of algorithms have been proposed for digital computation of the discrete FrFT. The most accurate and robust methods are based on Gauss–Hermite eigenfunctions, e.g. [13], and are computationally intensive, $\mathcal{O}(N^2)$, in comparison to a fast Fourier transform $\mathcal{O}(N\log(N))$. Although currently no fast implementation for exact calculation of the discrete FrFT exists, a fast implementation which calculates an approximation to the FrFT has been proposed [14].

2.2. Short-term fractional fourier transform (STFrFT)

Capus and Brown developed a STFrFT approach which is calculated in much the same way as the classical STFT [10]. The signal is windowed into discrete intervals in time which are each transformed in turn, at the appropriate fractional order α which yields the maximum response, to obtain the local power spectrum of the signal as a function of time. Rescaling from fractional to frequency domains can be achieved using Eq. (3). Capus and Brown demonstrated that a spectrogram based on a short-term fractional Fourier transform (STFrFT) analysis offers improved time–frequency resolution of chirp signals over conventional STFT analysis [10].

If the signal properties are known in advance, the value of α may be predetermined. Alternatively, if the signal properties are unknown α can be tuned with reference to the entire signal (globally optimized), or be selected independently for each window of analysis (locally optimized). The tuning process involves performing a FrFT over a range of fractional orders and selecting the order which best matches the chirp rate. Capus and Brown choose the α which yields the maximum amplitude response, although we have found that identifying the α which yields the greatest kurtosis also offers favorable results. The kurtosis, k, of a distribution is given by

$$k = \frac{\mu_4}{\mu_2^2}, \quad (4)$$

where μ_i denotes the ith central moment. Although tuning α to its optimum value may be quite computationally intensive, such a method has the advantage of requiring no *a priori* knowledge of the signal and produces the resolution needed to accurately measure the instantaneous frequency of a signal. However, the described methods of analysis are still not ideal for the analysis of signals which contain multiple chirps of different chirp rates, since any one value of α can only be optimal for one signal component, and will be sub-optimal for any other signal components which possess different chirp rates. Hence, these methods are not well suited to signals containing multiple non-related chirps.

This limitation means that this approach, for example, is not always ideally suited to radar or sonar tracking of multiple targets, or, as we will later demonstrate, to the analysis of bat echolocation signals.

2.3. Component optimized STFrFT (COSTFrFT)

In the STFrFT approach proposed here, the value of α is optimized for each signal component, whether stationary or moving. This is achieved through a process of iterative filtering in fractional Fourier domains in order to identify the optimum value of α for each component individually. FrFTs are performed over a range of orders α in a short-term analysis window of the signal. The α giving the maximum kurtosis is located, α_m, which identifies the chirp rate of the strongest component in the signal (Eq. (2)). The $F_{\alpha_m}(x)$ giving the maximum amplitude response, $F_{\alpha_m}(x_m)$, then identifies the center frequency of this component within the analysis window. The component

is recorded, then filtered out by applying a window, $W(x)$, which sets $F_{\alpha_m}(x_m)$ and the neighboring n values of x on either side of x_m to zero:

$$W(x)=\begin{cases}0, & (x_m-n)\le x\le(x_m+n),\\ 1 & \text{otherwise.}\end{cases} \tag{5}$$

We choose a fixed value of n, although one could base n on the measured width of the maximum. The original signal with this component removed is then constructed by performing the inverse FrFT (e.g. [15]). This entire process is iterated until no more signal components above a certain pre-determined amplitude threshold are detected. We choose a threshold which is 1% the amplitude of the strongest component. However, the pre-determined amplitude threshold could, for example, be a detection threshold based on a system's receiver operating characteristics, in order to provide a certain probability of detection for a given false alarm rate [6]. Alternatively, the number of iterations may be set to a fixed value if, for example, the number of signal components are known in advance.

The output is then constructed based on each of the stored $F_{\alpha_m}(x_m)$ components, projected onto the frequency domain. This process is repeated for each short-term analysis window to generate a new time–frequency representation analogous to a spectrogram. In effect, this processing scheme is equivalent to the STFT, but in each window of analysis each identified chirp is demodulated to its center frequency. It is noted that the approach to filtering in the fractional domain described in Eq. (5) works only for analytical signals, where the negative frequency content is minimized. Thus, for application to real signals, a Hilbert transform is applied first in order to construct the analytical signal.

We use a FrFT routine based on [14] to identify rapidly the optimal α during each iteration. However, due to its approximative nature, this routine suffers from poor accuracy over multiple iterations of our processing scheme. We therefore use the slower but more accurate FrFT routine based on [13] to recalculate $F_{\alpha_m}(x)$ and to then perform the inverse FrFT during each iteration. However, one of the reviewers has pointed out that the fast approximation to the discrete FrFT introduced by Pei and Ding [16] could be used to carry out both steps of the iteration since it is both fast and highly accurate and may therefore be more suitable.

It is noted that this processing scheme is best suited to the analysis of signals containing chirp-based components (which can be described as having a time-varying instantaneous frequency), such as bat echolocation calls, and is less suited to signals containing only band-limited noise or impulsive events such as airguns [17].

3. Results

We apply our component optimized STFrFT processing method to two signal types and compare its performance against other processing methods including the STFT, Wigner–Ville (W–V), pseudo Wigner–Ville (PW–V), smoothed pseudo Wigner–Ville (SPW–V), and a locally optimized STFrFT (LOSTFrFT, see Section 2.2). Routines for the W-V, PW-V and SPW-V were obtained from [18].

3.1. Multi-component chirp signal

The first example has been constructed to test the time and frequency localization ability of each method against a multi-component chirp signal; the signal consists of a linear and a weaker quadratic chirp and is of the form

$$s=e^{\mathrm{i}(at^2+bt)}+0.8e^{\mathrm{i}(xt^3+yt^2+zt)}, \tag{6}$$

where $a=-1\times10^7$, $b=7.1\times10^4$, $x=9.362\times10^9$, $y=4.59\times10^7$ and $z=-1.550\times10^4$. The sampling period is 7×10^{-6} s and the amplitude of the signal is tapered for $t<1.5\times10^{-3}$ s and $t>3.5\times10^{-3}$ s. Here, a 163 point Hann time-window with 162 point overlap was used for the STFT and the two STFrFT methods. In the STFrFT methods, the tuning process to find the optimal α was conducted in steps of $\Delta\alpha=0.01$ between $\alpha=0.75$ and 1.25. Additionally, a filter width of $n=2$ was used for the COSTFrFT and the iteration process was terminated immediately after finding the two strongest components. A 163 point Hann time-window was also used for the PW-V and SPW-V methods, with a 71 point Hann frequency-window applied to the latter. The various time–frequency representations of the signal are shown in Fig. 1. Each image has been normalized with respect to its maximum amplitude. In order to quantify the time–frequency resolution of each representation, we define the parameter R, which is simply the ratio of pixels that form the image which have an amplitude $>10\%$ of the maximum pixel amplitude. Although our choice of a 10% threshold is somewhat arbitrary, changing its value was found to have little impact on the results trend. A large value of R indicates that the signal energy is distributed over a wide area of time–frequency space, thus indicating poor time–frequency localization. A low value of R indicates that the majority of the signal energy is constrained to a small area, thus indicating high time–frequency localization.

The STFT (Fig. 1a) is clearly not ideal for displaying the time–frequency behavior of this signal: both the time and frequency localization of the two chirps is poor and they cannot be separated when close in frequency. Here, the STFT method yields a value of $R=0.12$. Additionally, the relative amplitudes of the two components are difficult to interpret. The displayed amplitude of a chirp on a spectrogram at any given instant is proportional to its rate of change of frequency, $\mathrm{d}f/\mathrm{d}t$, in addition to its intensity. Hence the amplitude of the quadratic chirp appears to vary throughout its duration.

The Wigner method (Fig. 1b) yields $R=0.096$ and offers significant improvement in the localization of the linear chirp but provides a very poor time–frequency representation of the quadratic chirp. Indeed Wigner distributions are not well suited to the analysis of non-linear chirps. Additionally, the relative amplitudes of the two components are not well represented. Furthermore, the spectrogram suffers from severe cross-term interference which is due to the inherent non-linearity of the

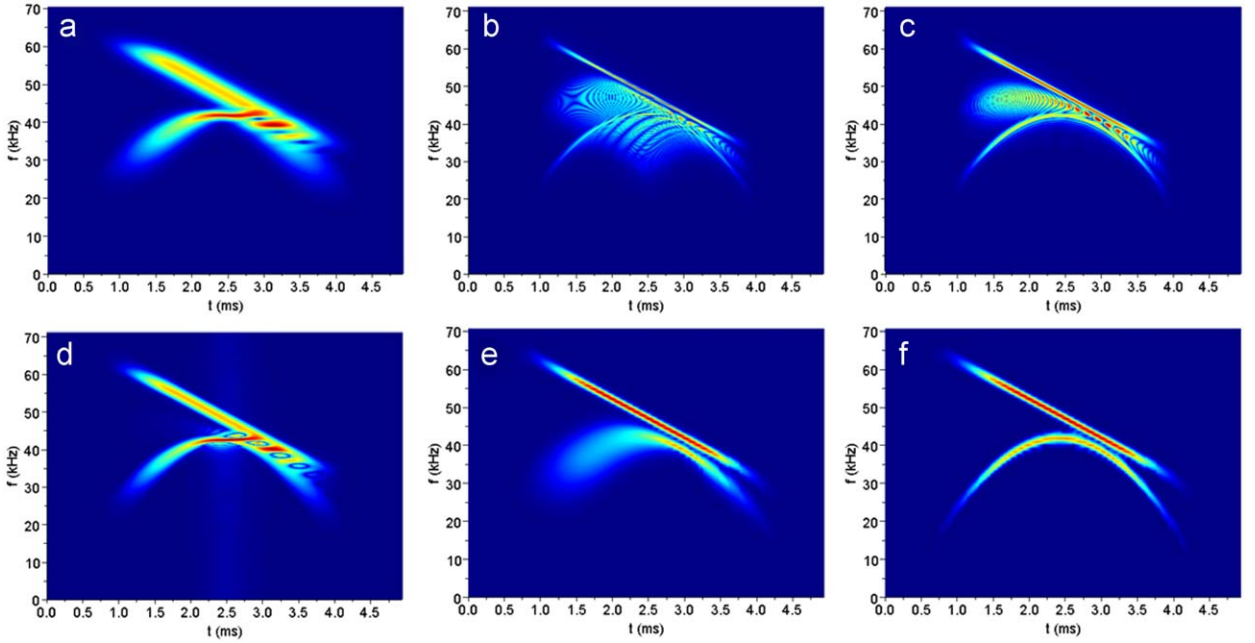

Fig. 1. Time–frequency representation of the multi-component chirp signal: STFT (a); W–V (b); PW–V (c); PSW–V (d); LOSTFrFT (e); and COSTFrFT (f).

Wigner transform. This results in the false appearance of signal energy at frequencies and times where no energy exists.

The PW–V method (Fig. 1c) offers a high degree of time and frequency localization for both the linear and quadratic chirps, yielding $R=0.078$, but this decomposition again suffers from cross-term interference resulting in artifacts at halfway in frequency between the two chirps. However, the decomposition does represent well the relative amplitudes of the two chirps. Cross-term interference effects are substantially reduced using the SPW-V scheme (Fig. 1d), resulting in a slightly improved resolution measure of $R=0.077$. However, the resulting spectrogram still does not offer sufficient time–frequency resolution to separate the two chirps when close in frequency, e.g. at $t=3.0$ ms, and captures poorly their relative amplitudes.

A locally optimized STFrFT [10] offers sufficient resolution to clearly identify that the two chirps do not cross over in frequency (Fig. 1e). However, this method is only optimized for the strongest identified component in the signal—which is the linear chirp in this example. The quadratic component therefore appears highly localized when its chirp rate is similar to that of the more energetic linear chirp, e.g. at $t=3.0$ ms, but has poor localization when its chirp rate differs from that of the linear chirp, e.g. when $t=2.0$ ms, thus yielding an overall resolution of $R=0.091$. The COSTFrFT (Fig. 1f), however, uses the optimal fractional order for each identified component providing high time and frequency localization ($R=0.055$) of both the quadratic and the linear chirp throughout their duration and without any cross-term artifacts. Furthermore, the relative amplitudes of the two components are well represented throughout their duration using the COSTFrFT approach. Hence, of the tested time–frequency representations, the COSTFrFT approach provides the most accurate time–frequency representation of this signal.

3.2. Bat echolocation signal

We now apply the processing schemes to a recording of a bat echolocation signal. This signal has a sampling period of 7×10^{-6} s and contains four approximately harmonically spaced and slightly non-linear chirps.

Fig. 2 shows the time–frequency representation of the signal using the various methods. Here, a 127 point Hann time-window with 126 point overlap was used for the STFT and the two STFrFT methods. In the STFrFT methods, the tuning process to find the optimal α was conducted in steps of $\Delta\alpha=0.005$ between $\alpha=0.7$ and 1.0. A filter width of $n=6$ was used for the COSTFrFT. A 127 point Hann time-window was also used for the PW-V and SPW-V methods, with a 65 point Hann frequency-window applied to the latter.

The STFT (Fig. 2a) yields a spectrogram which again offers poor time–frequency localization, with $R=0.17$. The W-V and PW-V methods (Figs. 2b and c) suffer from severe cross-term interference problems which results in the appearance of unwanted terms, which is reflected by a poor resolution measure of $R=0.19$ for both methods. The SPW-V method, Fig. 2d, offers an improvement over the previous techniques, with $R=0.12$. The problems of optimization are quite apparent for the LOSTFrFT shown in Fig. 2e. Although this method clearly offers better resolution than the previous techniques with $R=0.067$, a slight discontinuity is apparent at $t=2.0$ and 3.3 ms; this is a result of the optimization being based on the strongest identified component in each window of

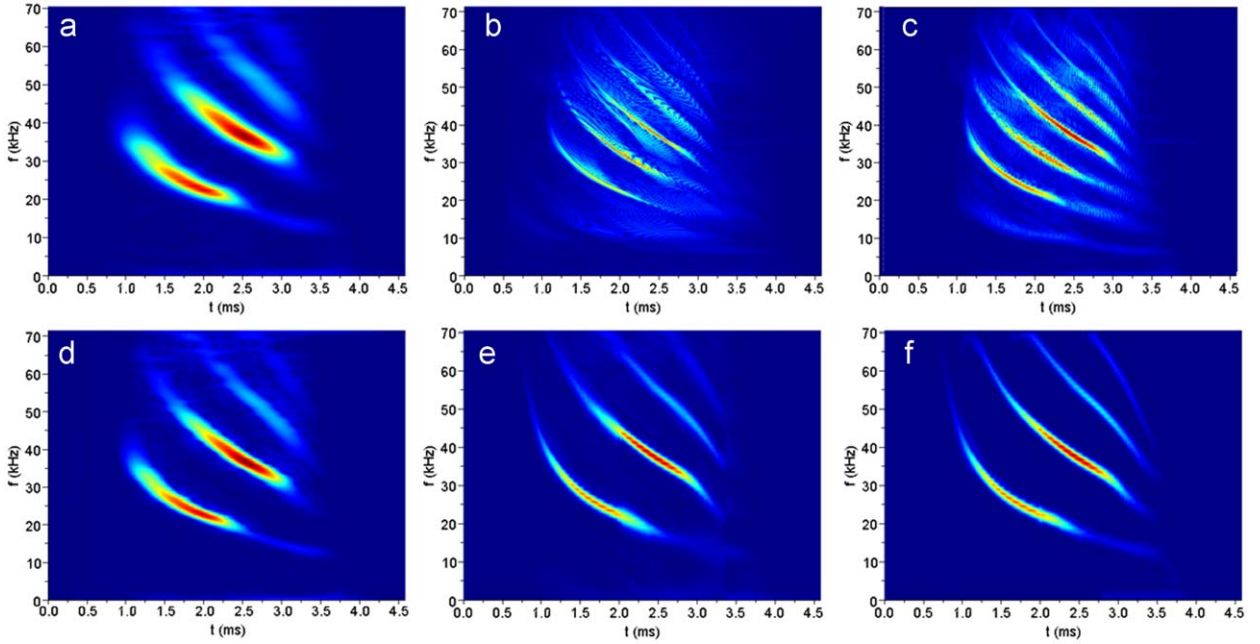

Fig. 2. Time–frequency representation of the bat echolocation signal: STFT (a); W–V (b); PW–V (c); PSW–V (d); LOSTFrFT (e); and COSTFrFT (f).

analysis, which changes during the signal. The COSTFrFT, Fig. 2f, does not suffer from this problem since it uses the optimal value of α independently for each detected signal component, providing a resolution measure of $R = 0.059$. The resulting spectrogram provides the most accurate representation of the echolocation signal, revealing a more detailed picture of its time–frequency characteristics. It is now clear that the echolocation signal is composed of a set of four time-delayed, amplitude modulated non-linear chirps. The variation between the components could help individual bats to distinguish their own echolocation signal from those of other bats of the same species. However, more research is required to support this hypothesis and the COSTFrFT is well suited to such a task.

3.3. Computational costs

Table 1 shows the relative computation time for calculation of each spectrogram shown in Fig. 1. These times are based on the average calculation time on the author's Dell Latitude D610 laptop running Matlab version R2008b on Microsoft Windows XP. Clearly, in this example, the COSTFrFT is by far the most computationally demanding method. For a signal containing N identifiable components, the COSTFrFT scheme requires a factor of approximately N more computations as the locally optimized STFrFT, which in turn requires approximately M times as many computations as a similar STFT, where M is the number of values of α over which the tuning process is conducted. Hence, the computation time for the COSTFrFT is dependant upon the signal under analysis.

Table 1
Computation time for calculation of spectrograms shown in Fig. 1.

Method	Computation time (s)
STFT	0.73
W–V	0.31
PW–V	0.28
SPW–V	4.07
LOSTFrFT	21.1
COSTFrFT	41.3

The computation times for the STFrFT methods are proportional to the number of values of α over which the tuning process is conducted. Thus, increasing the step size $\Delta\alpha$ over which the tuning process is conducted will reduce the computation time. However, increasing $\Delta\alpha$ will also tend to reduce the resolution of the resulting spectrogram. This is demonstrated in Fig. 3 which shows how R for the COSTFrFT varies with $\Delta\alpha$ for the artificial signal shown in Fig. 1. The smaller the step size in α, the greater the chance that the chirp rate of each signal component is always closely matched to one of the tested values of α. Therefore, a balance is required between a large $\Delta\alpha$ for computational efficiency and a small $\Delta\alpha$ to provide the desired level of resolution.

4. Conclusion

A method for producing a highly resolved spectrogram based on optimizing the fractional order for each identified component within a signal has been presented. We have demonstrated quantitatively that our relatively simple processing scheme offers spectrograms with improved time–frequency localization of signals contain-

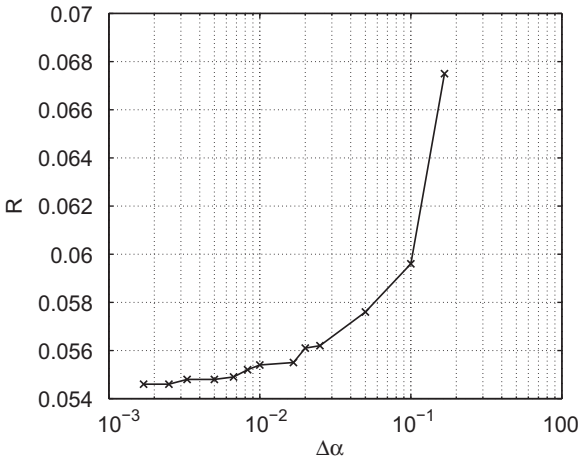

Fig. 3. Influence of step size $\Delta\alpha$ on COSTFrFT spectrogram resolution.

ing multiple non-linear chirps in comparison to several other established methods. In particular, the proposed processing scheme is well suited to the analysis of bat and marine mammal echolocation signals which are often comprised of multiple non-linear chirps. However, the improved resolution does come at a cost—typically at least an order of magnitude more computation is required for the COSTFrFT scheme than for the standard STFT. Thus, this scheme is most useful when other, less computationally intensive methods, do not provide the desired level of accuracy in representing the time–frequency behavior of a signal. In future research we aim to optimize the routine for identifying signal components with the goal of reducing the computational requirements.

Acknowledgment

We wish to thank A. Feng of the Beckman Institute for permission to use the bat echolocation signal.

References

[1] B.G. Ferguson, Time–frequency signal analysis of hydrophone data, IEEE J. Ocean. Eng. 21 (1996) 537–544.
[2] P.E. Howland, Target tracking using television-based bistatic radar, IEEE Proc. Radar Sonar Navig. 146 (1999) 166–174.
[3] K.J.W. Taylor, P.N. Burns, J.P. Woodcock, P.N.T. Wells, Blood flow in deep abdominal and pelvic vessels: ultrasonic pulsed-Doppler analysis, Radiology 154 (1985) 487–493.
[4] F. Hlawatsch, G.F. Boudreaux-Bartels, Linear and quadratic time–frequency signal representations, IEEE Signal Process. Mag. 9 (1992) 21–67.
[5] L. Cohen, Time–frequency distributions—a review, Proc. IEEE 77 (1989) 941–981.
[6] V.K. Madisetti, D.B. Williams, The Digital Signal Processing Handbook, CRC Press, Boca Raton, FL, 1998.
[7] S. Stenholm, The Wigner function: I. The physical interpretation, Eur. J. Phys. 1 (1980) 244–248.
[8] X.G. Xia, V.C. Chen, A quantitative SNR analysis for the pseudo Wigner–Ville distribution, IEEE Trans. Signal Process. 47 (1999) 2891–2894.
[9] S.A. Fulop, K. Fitz, Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications, J. Acoust. Soc. Am. 119 (2006) 360–371.
[10] C. Capus, K. Brown, Short-term fractional Fourier methods for the time–frequency representation of chirp signals, J. Acoust. Soc. Am. 113 (2003) 3253–3263.
[11] L.J. Stankovic, T. Alieva, M.J. Bastiaans, Time–frequency signal analysis based on the windowed fractional Fourier transform, Signal Process. 83 (2003) 2459–2468.
[12] L. Durak, A.K. Ozdemir, O. Arikan, Efficient computation of joint fractional Fourier domain signal representation, J. Opt. Soc. Am. A 25 (2008) 765–772.
[13] S. Pei, M. Yeh, C. Tseng, Digital fractional Fourier transform based on orthogonal projections, IEEE Trans. Signal Process. 47 (1999) 1335–1348.
[14] H.M. Ozaktas, M.A. Kutay, G. Bozdagi, Digital computation of the fractional Fourier transform, IEEE Trans. Signal Process. 44 (1996) 2141–2150.
[15] M. Bennet, S. McLaughlin, T. Anderson, W. McDicken, Filtering of chirped ultrasound echo signals with the fractional Fourier transform, in: IEEE Ultrasonics Symposium, vol. 3, 2004, pp. 2036–2040.
[16] S.C. Pei, J.J. Ding, Closed-form discrete fractional and affine Fourier transforms, IEEE Trans. Signal Process. 48 (2000) 1338–1353.
[17] J.E. Barger, W.R. Hamblen, The air gun impulsive underwater transducer, J. Acoust. Soc. Am. 68 (1980) 1038–1045.
[18] The Time–Frequency Toolbox, Available for download at: 〈http://tftb.nongnu.org/〉.