Time–Frequency Filtering Based on Spectrogram Zeros

Patrick Flandrin, *Fellow, IEEE*

***Abstract*—For a proper choice of the analysis window, a short-time Fourier transform is known to be completely characterized by its zeros, which coincide with those of the associated spectrogram. A simplified representation of the time-frequency structure of a signal can therefore be given by the Delaunay triangulation attached to spectrogram zeros. In the case of multicomponent nonstationary signals embedded in white Gaussian noise, it turns out that each time–frequency domain attached to a given component can be viewed as the union of adjacent Delaunay triangles whose edge length is an outlier as compared to the distribution in noise-only regions. Identifying such domains offers a new way of disentangling the different components in the time–frequency plane, as well as of reconstructing the corresponding waveforms.**

***Index Terms*—Delaunay triangulation, filtering, spectrogram, time–frequency analysis.**

I. Introduction

SPECTROGRAMS—i.e., squared magnitudes of Short-Time Fourier Transforms (STFTs)—are among the simplest and most natural tools for performing a time-frequency (TF) analysis of signals [8]. In the case of nonstationary signals with a limited number of components (e.g., AM-FM-type waveforms and/or impulse-like transients), spectrograms are relatively sparse representations, with a few energy ribbons localized along the TF trajectories of the different components. In the model-free situations we are interested in here, filtering those components is generally achieved by identifying non-parametrically the TF domains defining the ribbons, and then reconstructing waveforms by inverting the transform after masking.

This long-standing question has recently received a renewed interest, either because of the development of specific techniques such as “synchrosqueezing” [1], [6], [21] for which isolating domains of influence is a pre-requisite to reconstruction [19], or because of new proposals such as “contours” [18] for defining basins of attraction attached to components. In most cases, the rationale for identifying “signal regions” is based on local energy considerations and/or curves such as ridges [7] which are supposed to capture local energy concentration. In contrast with such approaches based on *large values* of the representation, we will propose here to make use of *zeros* as characteristic points. Two main features of the proposed approach are that (1) it allows for an *unsupervised disentangling* and identification of individual components (in contrast with simple schemes based on energy thresholding), and (2) its geometrical, non parametric rationale makes it *rotation-invariant* in the TF plane, with a similar ability to deal with impulse-like transients signals (with almost “vertical” TF signatures) and AM-FM-type waveforms (with almost “horizontal” ones).

The paper is organized as follows. Section II is devoted to STFT/spectrogram, with basics recalled in Section II-A and elementary facts on the white Gaussian noise case in Section II-B. Section III then discusses more specifically the role played by zeros in a STFT/spectrogram: Section III-A justifies the complete representation they offer for a proper choice of the short-time window, whereas Section III-B suggests the Delaunay triangulation based on zeros as a way of getting a simplified description. This paves the way for the new filtering approach that is discussed and illustrated in Section IV.

II. STFT and Spectrogram

A. Definitions and Basics

Given a signal $x(t)$ and a window $h(t)$, the STFT $F_x^{(h)}(t,\omega)$ is classically defined as the inner product between $x(t)$ and shifted versions (in time and frequency) of $h(t)$, i.e., as $F_x^{(h)}(t,\omega) = \langle x, \mathbf{T}_{t\omega} h\rangle$, where $\mathbf{T}_{t\omega}$ stands for some joint TF shift operator. For a sake of simplicity and symmetry, we fix here $\mathbf{T}_{t\omega}$ in reference to the *Weyl operator*, [4], [5], thus ending up with the explicit definition:

$$F_x^{(h)}(t,\omega) = \int_{-\infty}^{+\infty} x(s)\overline{h(s-t)} \exp\left\{-i\omega\left(s-\frac{t}{2}\right)\right\} ds. \tag{1}$$

The corresponding spectrogram simply follows as:

$$S_x^{(h)}(t,\omega) = \left|F_x^{(h)}(t,\omega)\right|^2. \tag{2}$$

It is well-known that $F_x^{(h)}(t,\omega)$ and $S_x^{(h)}(t,\omega)$ are not any 2D functions since that, by construction, they inherit some struc-

Manuscript received April 08, 2015; revised July 20, 2015; accepted July 26, 2015. Date of publication August 05, 2015; date of current version August 07, 2015. This work was supported by Agence Nationale de la Recherche under Grant ANR-13-BS03-0002-01 which is part of the ASTRES Project. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiao-Ping Zhang.

The author is with the Physics Department (UMR 5672 CNRS), Ecole Normale Supérieure de Lyon, 69364 Lyon Cedex 07 France (e-mail: patrick.flandrin@ens-lyon.fr).

Digital Object Identifier 10.1109/LSP.2015.2463093

ture from their definition as inner products. More precisely, the STFT (1) satisfies the reproducing identity:

$$F_x^{(h)}(t',\omega') = \int\int_{-\infty}^{+\infty} K(t',\omega';t,\omega) F_x^{(h)}(t,\omega)dt\frac{d\omega}{2\pi}, \quad (3)$$

in which the reproducing kernel $K(t',\omega';t,\omega)$ is (up to a multiplicative term) nothing but the STFT of the window:

$$K(t',\omega';t,\omega) = \frac{\exp\{i(\omega t' - \omega' t)/2\}}{\|h\|_2^2} F_h^{(h)}(t'-t,\omega'-\omega). \quad (4)$$

An equivalent formulation amounts to rewriting $F_h^{(h)}(t,\omega)$ as $A_h(\omega,t)$, with

$$A_x(\xi,\tau) = \int_{-\infty}^{+\infty} x\left(\theta+\frac{\tau}{2}\right)\overline{x\left(\theta-\frac{\tau}{2}\right)}\exp\{-i\xi\theta\}d\theta \quad (5)$$

the so-called *ambiguity function*, [8]. Doing so, it is easy to justify that any STFT (or spectrogram) has necessarily some local redundancy since the reproducing kernel (4) cannot be arbitrarily peaked in both time and frequency. This follows from general uncertainty relations attached to ambiguity functions in terms of support and/or volume (see, e.g., [17] or [13] for precise formulations). The consequence of these inequalities is that the reproducing kernel (4) has necessarily some non-zero extension that controls the local redundancy of the STFT/spectrogram. In the particular case of the (unit energy) Gaussian window[1]

$$g(t) = \pi^{-1/4}\exp\{-t^2/2\}, \quad (6)$$

which is referred to as "circular" since

$$A_g(\xi,\tau) = \exp\{-(\xi^2+\tau^2)/4\}, \quad (7)$$

the reproducing kernel is maximally concentrated and defines an "influence domain" which is circular and whose radius is given by some effective area attached to the 2D Gaussian function of variance 2.

B. Spectrogram of White Gaussian Noise

Whereas the spectrogram is classically defined for finite energy, deterministic signals, it can also be used for the analysis of finite power, harmonizable random processes [8][2]. More specifically, we will restrict to the idealized case of zero-mean, analytic, white Gaussian noise $n(t)$ such that:

$$\mathbb{E}\{n(t)\overline{n(t')}\} = \gamma_0\delta(t-t'); \quad \mathbb{E}\{n(t)n(t')\} = 0. \quad (8)$$

It readily follows from (2) and (8) that the expected value of the spectrogram is in this case constant:

$$\mathbb{E}\left\{S_n^{(h)}(t,\omega)\right\} = \gamma_0^2, \quad (9)$$

whereas the covariance between spectrogram values at two different locations in the TF plane only depends on the corresponding lags in both time and frequency, according to the relation:

$$\mathrm{cov}\left\{S_n^{(h)}(t,\omega), S_n^{(h)}(t',\omega')\right\} = \gamma_0^2 S_h^{(h)}(t'-t,\omega-\omega'). \quad (10)$$

In the specific case of the circular Gaussian window (6), this covariance takes on the simple form

$$\mathrm{cov}\left\{S_n^{(g)}(t,\omega), S_n^{(g)}(t',\omega')\right\} = \gamma_0^2\exp\{-d^2((t,\omega),(t',\omega'))/2\}, \quad (11)$$

where $d((t,\omega),(t',\omega')) = \sqrt{(t-t')^2+(\omega-\omega')^2}$ measures the Euclidian distance in the plane between the two considered points.

As a function of this only distance, the spectrogram of white Gaussian noise can then be considered as a second-order homogeneous (or stationary) field. This homogeneity property carries over to characteristic points of the surface (such as extrema, be they local maxima or zeros). However, due to the reproducing kernel structure recalled above, the distribution of those characteristics points is expected to be constrained as well.

III. Spectrogram Zeros

A. The Bargmann Connection

Time and frequency are usually considered either independently or jointly, but it might be interesting to see them as coordinates of a complex-valued variable, thus identifying the TF plane with the complex plane. Doing so by introducing $z = \omega + it$, a direct calculation shows that, when evaluated with the circular Gaussian window $g(t)$ defined in (6), the STFT (1) can be re-written as:

$$F_x^{(g)}(t,\omega) = \exp\{-|z|^2/4\}\mathcal{F}_x(z), \quad (12)$$

where

$$\mathcal{F}_x(z) = \int_{-\infty}^{+\infty} A(z,s)x(s)ds \quad (13)$$

and

$$A(z,s) := \pi^{-\frac{1}{4}}\exp\{-s^2/2 - isz + z^2/4\}. \quad (14)$$

This corresponds to the Bargmann factorization of the STFT, with (13) the Bargmann transform [2], whose kernel is given by (14). One interest of such a companion formulation for the STFT is that (13) is an entire function, with consequences on the structure of the STFT and the associated spectrogram. More specifically, since the circular Gaussian window (6) is normalized so as to be of unit energy, this immediately results in the upper bound $|F_x^{(g)}(t,\omega)| \leq \|x\|$. Together with the factorization (12), this leads to

$$|\mathcal{F}_x(z)| \leq \|x\|\exp\{|z|^2/4\}, \quad (15)$$

[1]In the Physics literature (see, e.g., [15]), the corresponding spectrogram is referred to as the "Husimi distribution function" [14].

[2]We will here formally apply the definition (2) to finite duration realizations of such stochastic processes, looking at statistical properties of the corresponding spectrogram characteristics.

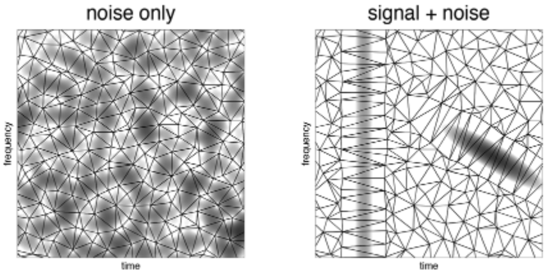

Fig. 1. **Spectrogram and Delaunay triangulation**. Left: In the case of white Gaussian noise, the Delaunay triangulation constructed upon the zeros of the spectrogram reveals an homogeneous distribution of random triangles. Right: when a signal (namely, an impulse followed by a chirped Gaussian in this example) is superimposed to noise, the distribution of the Delaunay triangles remains unaffected in noise-only regions, whereas the "signal domains" are characterized by more elongated triangles, with some edges longer than expected in the noise-only case.

i.e., to the fact that $\mathcal{F}_x(z)$ is an entire function of order 2 [3]. As a consequence, it admits a Weierstrass-Hadamard factorization of the form [3], [12], [15], [22]

$$\mathcal{F}_x(z) \propto \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right) \exp\left\{\frac{z}{z_n} + \frac{1}{2}\left(\frac{z}{z_n}\right)^2\right\}, \quad (16)$$

where the variables $z_n = \omega_n + it_n$ stand for the (infinitely many) zeros of the Bargmann transform which, by construction, also correspond to the zeros of the STFT and of the spectrogram. Although (16) is unlikely to be used as such for a possible reconstruction, its meaning is that the Bargmann transform (and, hence, the associated STFT/spectrogram) is completely characterized by the distribution of its zeros.

B. Delaunay Triangulation

Since zeros completely characterize a STFT, it is expected that their distribution in the TF plane—the so-called "stellar representation" in Quantum Mechanics [16] (see also [12] for a related TF perspective)—evidences distinctive properties attached to the nature of the analyzed signal. In this respect, we can therefore get a simplified, geometrical description of the TF structure of a signal by looking at diagrams connecting STFT/spectrogram zeros, the simplest one being the *Delaunay triangulation*, [20].

An example of a Delaunay triangulation attached to the collection of STFT/spectrogram zeros in the case of white Gaussian noise is given in Fig. 1 (left diagram). Since the stationarity of the analyzed white Gaussian noise results in the homogeneity of the 2D random field defined by the STFT/spectrogram (see Section II-B), the distribution of zeros is itself homogeneous all over the plane, a situation that is expected to be broken whenever some signal—with a coherent TF structure, such as a frequency modulation—happens to be superimposed. As evidenced in the same Fig. 1 (right diagram), this is exactly what happens: when a signal (impulse + AM-FM chirp) is added to the noise of the left diagram, the noise-only regions remain unaffected whereas the "signal domains" are characterized not only by large spectrogram values but also by Delaunay triangles that are more elongated and with longer edges than in noise-only regions.

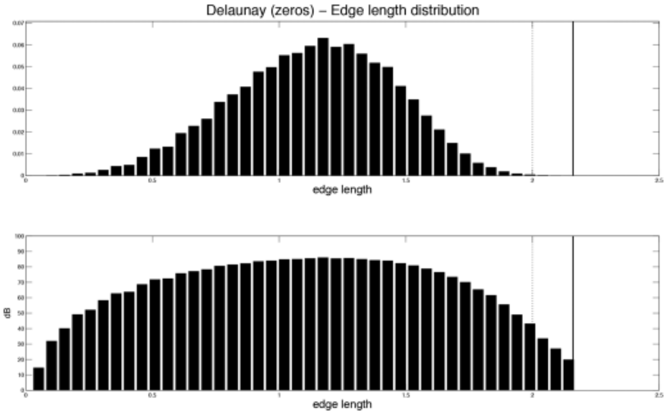

Fig. 2. **Delaunay triangulation—Distribution of edge lengths**. In the case of white Gaussian noise, the distribution of edge lengths in Delaunay triangles constructed upon spectrogram zeros (top: linear scale; bottom: logarithmic scale) is essentially bounded above by a maximum length $L_{max} \sim 2.2$ (full line). Moreover, the probability that the edge length exceeds the value 2 (dotted line) is about 10^{-3}.

IV. Time-Frequency Filtering

A. Rationale

Both the theoretical considerations of the previous sections and the evidences of Fig. 1 suggest that signal domains can be identified by looking at Delaunay triangles that depart from the expected behavior attached to noise, thus calling for a characterization of this reference situation. In this respect, Fig. 2 displays the distribution of edge lengths of Delaunay triangles constructed upon STFT/spectrogram zeros in the nominal case of white Gaussian noise. What is evidenced is that such a length is essentially bounded above by a maximum value $L_{max} \sim 2.2$, with a very low probability to exceed, e.g., 2 (referring as $|e_{mn}|$ the distance $d(z_m, z_n)$ between any two zeros z_m and z_n, a numerical evaluation shows that $\text{Prob}\{|e_{mn}| > 2\} \sim 10^{-3}$). Selecting Delaunay triangles on the basis of thresholding their maximum edge length is therefore a simple way of identifying elementary signal domains whose concatenation defines supports—delineated by zeros—for TF 1/0 masks to be applied to STFT prior reconstruction of the corresponding signal components.

B. Algorithm

Based on the elements obtained above, the TF filtering algorithm is quite straightforward and can be summarized as follows:

1. Perform Delaunay triangulation over STFT zeros z_m;
2. Identify outlier edges such that $|e_{mn}| = d(z_m, z_n) > \eta$, with the threshold chosen typically as $\eta \approx 2$;
3. Keep triangles with at least one outlier edge;
4. Group adjacent such triangles in connected, disjoint domains $\mathcal{D}_j$;
5. Multiply STFT with labeled 1/0 masks $\mathbf{1}_{\mathcal{D}_j}(t, \omega)$;
6. Reconstruct the disentangled components, domain by domain, by using the standard formula:

$$x_j(t) = \int\int_{(s,\omega)\in\mathcal{D}_j} F_x^{(h)}(s, \omega) \left(\mathbf{T}_{s\omega} h\right)(t) ds \frac{d\omega}{2\pi}. \quad (17)$$

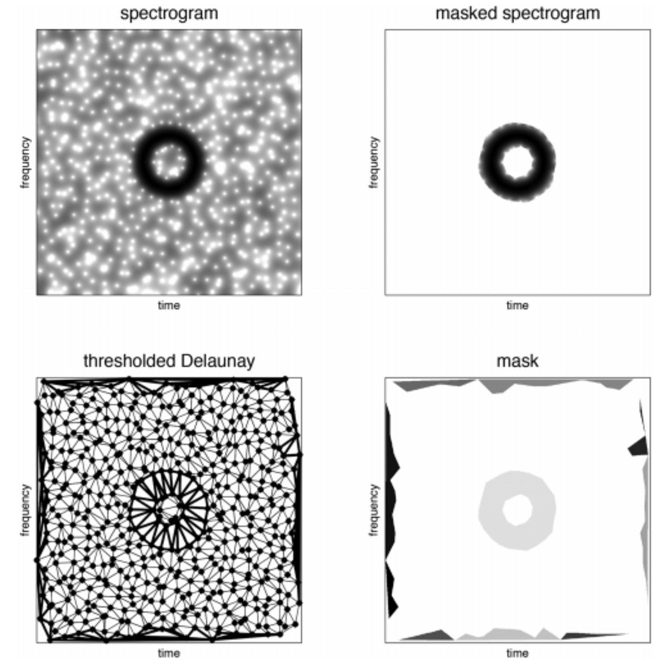

Fig. 3. **Time-frequency filtering—Synthetic signal example (1/2)**. Top left: spectrogram of a Hermite function embedded in white Gaussian noise (SNR $= 10$ dB). Bottom left: Delaunay triangulation constructed on the zeros of the spectrogram, with outlier edges (see text) highlighted as thicker lines. Bottom right: time-frequency domains obtained by concatenating adjacent Delaunay triangles with outlier edges, each domain being labeled by a different grey level. Top right: masked spectrogram when retaining as domain the central ribbon.

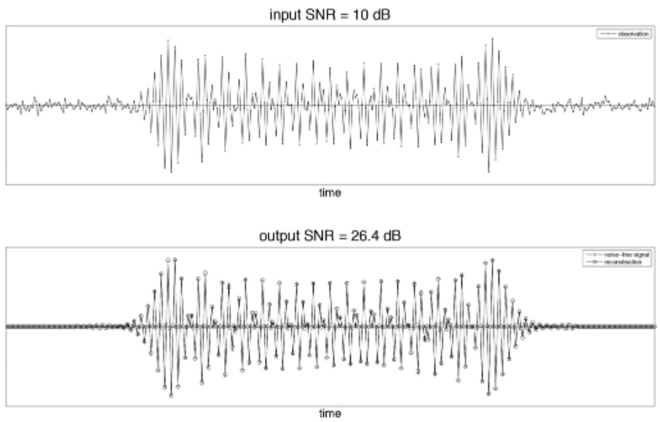

Fig. 4. **Time-frequency filtering—Synthetic signal example (2/2)**. Top: noisy observation. Bottom: reconstructed waveform obtained by inverting the masked STFT of Fig. 3 (top right), together with the noise-free Hermite function for a sake of comparison.

C. Examples

The first example consists in a Hermite function (of order 17), whose TF "trajectory" is known to be a circle [9] (a situation that cannot be easily parameterized within the frameworks of AM-FM signals as considered, e.g., in [7], [21] or [23]). The overall filtering procedure (triangulation, selection of outliers, grouping and masking) is summarized in Fig. 3, with the corresponding reconstruction result in Fig. 4, evidencing a gain of about 16 dB.

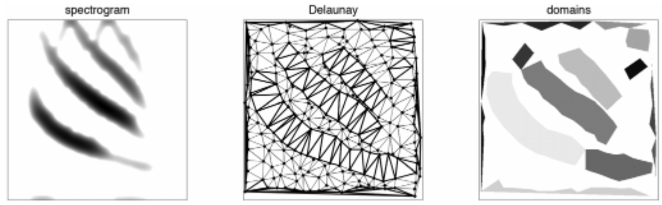

Fig. 5. **Time-frequency filtering—Real data example (1/2)**. Left: spectrogram of the benchmark "bat signal". Middle: Delaunay triangulation constructed on the zeros of the spectrogram, with outlier edges (see text) highlighted as thicker lines. Right: time-frequency domains obtained by concatenating adjacent Delaunay triangles with outlier edges, each domain being labeled by a different grey level.

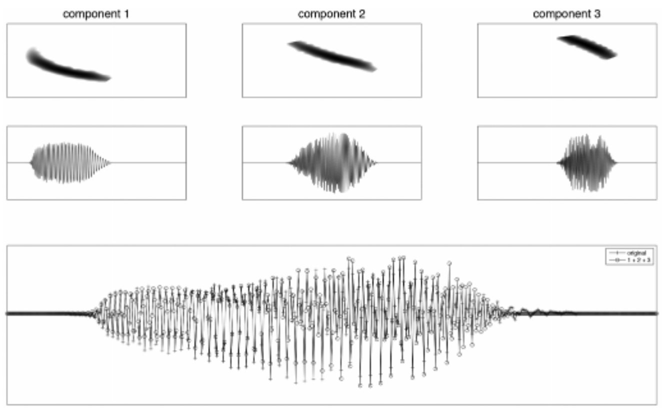

Fig. 6. **Time-frequency filtering—Real data example (2/2)**. Top row: masked spectrograms of the 3 main components identified in Fig. 5. Middle row: the corresponding waveforms obtained by inverting the respective masked STFTs. Bottom row: superimposition of the above 3 components, together with the original signal for a sake of comparison.

The second example corresponds to the classical benchmark "bat signal"[3], whose spectrogram and Delaunay selection of the domains corresponding to the different components are presented in Fig. 5, with individual reconstructions of the 3 main ones plotted in Fig. 6, together with their recombination to be compared to the complete waveform.

V. Concluding Remarks

A new approach to TF filtering has been proposed, based on identifying spectrogram zeros rather than thresholding energy levels. The advantage is expected to be threefold: (1) anchoring signal domains to zeros allows them to be maximally large and capture all of the component information; (2) grouping Delaunay triangles prior filtering permits an unsupervised disentanglement of signals into labeled components; (3) being purely geometrical, the approach is naturally rotation-invariant in the TF plane, with an equal ability to deal with TF trajectories of any orientation (be they almost "vertical" for impulse-like transients or "horizontal" for AM-FM-type waveforms). Only the rationale of the approach has been presented here, with examples supporting its effectiveness. Further studies will be needed, e.g., to tune the edge length threshold η which controls the detection/false alarm trade-off, as well as to better assess its performance and compare with competing methods.

[3]Thanks to C. Condon, K. White and A. Feng of the Beckman Institute of the University of Illinois for the bat data and for permission to use it in this paper.

REFERENCES

[1] F. Auger, P. Flandrin, Y. T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H. T. Wu, "Time-frequency reassignment2 and synchrosqueezing: An overview," *Proc. IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 32–41, Nov. 2013.

[2] V. Bargmann, "On a Hilbert space of analytic functions and an associated integral transform," *Commun. Pure Appl. Math.*, vol. 14, pp. 187–214, 1961.

[3] R. P. Boas, *Entire Functions*. New York, NY, USA: Academic, 1954.

[4] E. Chassande-Mottin, I. Daubechies, F. Auger, and P. Flandrin, "Differential reassignment," *IEEE Signal Process. Lett.*, vol. 4, no. 10, pp. 293–294, Oct. 1997.

[5] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.

[6] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Appl. and Comp. Harm. Anal.*, vol. 30, no. 1, pp. 243–261, 2011.

[7] N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrésani, "Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 644–664, Mar. 1992.

[8] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. New York, NY, USA: Academic, 1999.

[9] P. Flandrin, "A note on reassigned Gabor spectrograms of Hermite functions," *J. Fourier Anal. Appl.*, vol. 19, no. 2, pp. 285–295, 2012.

[10] P. Flandrin, F. Auger, and E. Chassande-Mottin, "Time-frequency reassignment? from principles to algorithms," in *Applications in Time-Frequency Signal Processing, A. Papandreou-Suppappola*. Boca Raton, FL, USA: CRC, 2003, ch. 5, pp. 179–203.

[11] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pp. 429–457, 1946.

[12] T. J. Gardner and M. O. Magnasco, "Sparse time-frequency representations," in *Proc. Nat. Acad. Sci.*, 2006, vol. 103, no. 16, pp. 6094–6099.

[13] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Boston, MA, USA: Birkhäuser, 2011.

[14] K. Husimi, "Some formal properties of the density matrix," in *Proc. Phys. Math. Soc. Jpn*, 1940, vol. 22, pp. 264–314.

[15] H. J. Korsch, C. Muller, and H. Wiescher, "On the zeros of the Husimi distribution," *J. Phys. A: Math. Gen.*, vol. 30, pp. L677–L684, 1997.

[16] P. Lebœuf and A. Voros, "Chaos-revealing multiplicative representation of quantum eigenstates," *J. Phys. A Math. Gen.*, vol. 23, pp. 1765–1774, 1990.

[17] E. H. Lieb, "Integral bounds for radar ambiguity functions and Wigner distributions," *J. Math. Phys.*, vol. 31, pp. 594–599, 1990.

[18] Y. Lim, B. G. Shinn-Cunningham, and T. J. Gardner, "Sparse contour representations of sound," *IEEE Signal Proc. Lett.*, vol. 19, no. 10, pp. 684–687, Oct. 2012.

[19] S. Meignen, T. Oberlin, and S. McLaughlin, "On the mode synthesis in the synchrosqueezing method," in *Proc. Eur. Signal Process. Conf. EUSIPCO-12*, Bucharest, Romania, 2012.

[20] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations—Concepts and Applications of Voronoi Diagrams*, 2nd ed. New York, NY, USA: Wiley, 2000.

[21] G. Thakur and H.-T. Wu, "Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples," *SIAM J. Math. Anal.*, vol. 43, pp. 2078–2095, 2011.

[22] M. Toda, "Phase retrieval problem in quantum chaos and its relation to the origin of irreversibility I," *Physica D*, vol. 59, pp. 121–141, 1992.

[23] Y. Yang, X. Dong, Z. Peng, and W. Zhang, "Component extraction for non-stationary multi-component signal using parameterized de-chirping and band-pass filter," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1373–1377, Sep. 2015.