Uncertainty and Spectrogram Geometry

Patrick Flandrin, E. Chassande-Mottin, François Auger

UNCERTAINTY AND SPECTROGRAM GEOMETRY

Patrick Flandrin[1] *Éric Chassande-Mottin*[2] *François Auger*[3]

[1] Laboratoire de Physique de l'École Normale Supérieure de Lyon, CNRS and Université de Lyon, France

[2] Laboratoire AstroParticule et Cosmologie (APC), Université Paris Diderot, CNRS, CEA, Observatoire de Paris, France

[3] LUNAM Université, IREENA, Saint-Nazaire, France.

patrick.flandrin@ens-lyon.fr, ecm@apc.univ-paris7.fr, francois.auger@univ-nantes.fr

ABSTRACT

Ultimate possibilities of localization for time-frequency representations are first reviewed from a joint perspective, evidencing that Heisenberg-type pointwise limits are not exclusive of sharp localization along trajectories in the plane. Spectrogram reassignment offers such a possibility and, in order to revisit its connection with uncertainty, geometrical properties of spectrograms are statistically investigated in the generic case of white Gaussian noise. Based on Voronoi tessellations and Delaunay triangulations attached to extrema, it is shown that, in a first approximation, local energy "patches" are distributed according to a randomized hexagonal lattice with a typical scale within a factor of a few that of minimum uncertainty Gabor logons.

Index Terms— Time-frequency, uncertainty, spectrogram, reassignment

1. FROM UNCERTAINTY...

1.1. From time and frequency to time-frequency

It is well-known that a square-integrable signal $x(t)$ cannot have its energy E_x arbitrarily localized in both time and frequency [1, 2]. The most common formulation of this limitation makes use of second-order measures for the spreadings in the two domains, namely:

$$\Delta_t^2(x) := \frac{1}{E_x} \int t^2 \, |x(t)|^2 \, dt; \tag{1}$$

$$\Delta_\omega^2(X) := \frac{1}{E_x} \int \omega^2 \, |X(\omega)|^2 \, \frac{d\omega}{2\pi}, \tag{2}$$

where $X(\omega)$ stands for the spectrum of $x(t)$.

Given such measures (which assume for simplicity, but with no loss of generality, that the individual densities $|x(t)|^2$ and $|X(\omega)|^2$ are centered), it readily follows that:

$$\Delta_t(x) \, \Delta_\omega(X) \geq \frac{1}{2}, \tag{3}$$

with equality if and only if the signal is Gaussian-shaped:

$$x(t) = C \, e^{\alpha \, t^2}, \; \alpha < 0. \tag{4}$$

Since both variables of time t and (angular) frequency ω are simultaneously involved in the basic "uncertainty" relation (3), it has been proposed [2, 3] to revisit the very same limitation from a joint perspective, based on the use of time-frequency (TF) energy distributions. In this respect, the most natural and simple counterpart of (3) reads

$$\Delta_{t\omega}(C_x) := \frac{1}{E_x} \iint \left(\frac{t^2}{T^2} + T^2 \omega^2 \right) \, C_x(t, \omega; \varphi) \, dt \, \frac{d\omega}{2\pi}, \tag{5}$$

where T is some arbitrary time scale and $C_x(t, \omega; \varphi)$ stands for any element of Cohen's class [1], as parameterized by its kernel function $\varphi(\xi, \tau)$ in the (2D Fourier transformed) ambiguity domain. Whereas general results can be found, e.g., in [2], one can single out the two most significant special cases, corresponding respectively to the Wigner distribution $W_x(t, \omega)$ and the spectrogram $S_x^h(t, \omega)$ (with window $h(t)$). In the first case, one can show that, for any T,

$$\Delta_{t\omega}(W_x) \geq 1, \tag{6}$$

whereas, in the second one, the inequality reads

$$\Delta_{t\omega}(S_x^h) \geq 2. \tag{7}$$

In both situations, the lower bound is reached (as in the classical formulation) for Gaussian signals, with furthermore the need of a Gaussian window in the spectrogram case. Since it is well-known that a spectrogram results from the smoothing of the Wigner distribution of the signal by that of the window [1, 2], the doubling of the minimum uncertainty observed when passing from (6) to (7) can be simply interpreted as the summing up of the corresponding spreads.

1.2. Heisenberg refined

In its simplest form (3), the "Heisenberg" inequality refers to the minimum achievable uncertainty under the assumption that the time and frequency variables are uncoupled. Relaxing this assumption in the proof, one can end up with the refined "Schrödinger" inequality [4]

$$\Delta_t(x) \, \Delta_\omega(X) \geq \frac{1}{2} \sqrt{1 + c^2(x)} \tag{8}$$

which explicitly makes use of the covariance measure:

$$c(x) := \int t\,|x(t)|^2\,\frac{d}{dt}\arg x(t)\,dt. \qquad (9)$$

As explained in [1], the covariance (9) can be thought of as an average, with respect to the energy density $|x(t)|^2$, of the product between time and the "instantaneous frequency" $\omega_x(t)$ defined as the phase derivative of the signal:

$$c(x) = \langle t\,\omega_x(t)\rangle. \qquad (10)$$

If these two quantities are "independent", one expects, under the assumption of centered densities, that

$$c(x) = \langle t\,\omega_x(t)\rangle = \langle t\rangle\,\langle\omega_x(t)\rangle = \langle t\rangle\,\langle\omega\rangle = 0 \qquad (11)$$

and therefore the recovering of (3) from the vanishing of the covariance in (8). However, when this covariance is non-zero, the lower bound is increased, and its maximum value is reached when the coupling between t and $\omega_x(t)$ is maximized. It follows from the Cauchy-Schwarz inequality that this happens in the colinear case corresponding to a quadratic phase. This means that minimizers of the Schrödinger inequality (8) are of the form:

$$x(t) = e^{\alpha\,t^2+\beta\,t+\gamma},\ \mathrm{Re}\{\alpha\} < 0. \qquad (12)$$

In Physics terms, such waveforms are referred to as "squeezed states" (as opposed to the (Glauber) "coherent states" defined in (4)) [5] whereas, in Signal Processing, they correspond to "linear chirps" with a Gaussian envelope.

2. ...TO LOCALIZATION

2.1. Perfect chirp localization

In its most constrained form (3), the uncertainty relation forbids any *pointwise* localization of energy in both time and frequency, with the consequence that no signal can have its energy concentrated on a TF domain of arbitrarily small area. This however does not rule out other forms of localization along *trajectories* in the TF plane. This is illustrated symbolically in Fig. 1, where a TF localized linear chirp can be seen as the limit of the waveform (12) when $\mathrm{Re}\{\alpha\} \to 0_-$ and $\mathrm{Im}\{\alpha\} \neq 0$. Indeed, when expressed in TF terms, the existence of a strictly positive lower bound in uncertainty relations implies that any energy distribution necessarily extends over a TF domain with some minimum, non-zero area. Given this area, the shape of the corresponding domain can however be varied, thanks to possible covariances in the TF plane with respect to, e.g., scalings and rotations. Within this picture, an initial "circular" logon (left) can be either compressed or dilated in one variable (with a corresponding dilation or compression in the dual variable so as to keep the area unchanged). Such transformations can also be combined with rotations, ending up eventually with a linear localized structure (right) in the limit of an infinite flattening of the envelope. This illustration is not only symbolic, but can also correspond to actual distributions: this is the case with the Wigner distribution which takes on a 2D Gaussian form when applied to waveforms such as (12) and which—thanks to its covariance with respect to scalings and rotations—is known to perfectly localize along the straight line of instantaneous frequency in the limit case of a constant magnitude linear chirp [1, 2].

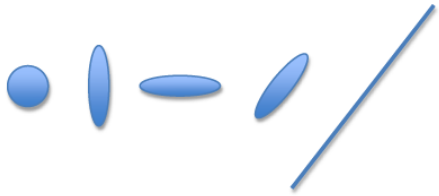

Fig. 1. *Uncertainty and time-frequency localization.*

2.2. Reassignment as a substitute

At first sight, the Wigner distribution may seem to be an adequate solution for achieving maximum localization in the TF plane, but a second thought highlights at least two limitations in the approach. First, localization can only be guaranteed for *linear* chirps and, second, the quadratic nature of the Wigner distribution creates interference patterns that confuse the picture as soon as more than one component is present at the same time [2]. The first limitation can be overcome to some extent by replacing the Wigner distribution by variants that are matched to specific forms of nonlinear chirps, but it can be shown [6] that localization still imposes the transform to be quadratic, thus leaving unchanged the interference terms issue. An efficient, yet approximate way out is however possible. This is based on the idea of *reassignment* [7], whose basics can be briefly recalled as follows in the case of the spectrogram (though the principle can be extended to more general distributions, see, e.g., [8, 9]).

The starting point of reassignment is to re-express a spectrogram, usually defined as the squared magnitude of a Short-Time Fourier Transform (STFT), as the 2D smoothing of the Wigner distribution of the signal by that of the window:

$$S_x^h(t,\omega) = \iint W_x(s,\xi)\,W_h(s-t,\xi-\omega)\,dt\,\frac{d\omega}{2\pi}. \qquad (13)$$

This allows for a simple interpretation: the value of a spectrogram at some given TF point (t,ω) results from the summing up of all local values of the Wigner distribution within a domain whose extension is essentially the Heisenberg cell of the window. Unless such values would be symmetrically distributed around it, the geometrical center of this cell has however no reason to be chosen as the locus where to

assign the integrated local energy. Indeed, a more meaningful location is the centroïd of the Wigner distribution values within the cell, and the purpose of reassignment is precisely to move each spectrogram value from the point (t,ω) where it has been computed to such a centroïd $(\hat{t}_x(t,\omega),\hat{\omega}_x(t,\omega))$:

$$\hat{S}_x^h(t,\omega) = \iint S_x^h(\tau,\xi)\delta(t-\hat{t}_x(\tau,\xi),\omega-\hat{\omega}_x(\tau,\xi))d\tau\frac{d\xi}{2\pi}. \quad (14)$$

From a practical point of view, the spectrogram with window $h(t)$ is classically computed as the squared magnitude of the corresponding Short-Time Fourier Transform (STFT) $F_x^h(t,\omega)$, and the identification of the centroïds coordinates can be efficiently achieved by supplementing this computation with that of two additional STFT's based on the companion windows $t.h(t)$ and $dh(t)/dt$ [8, 9].

2.3. Example and interpretation

In the following, we will restrict to Gaussian windows of the form:

$$h(t) = \pi^{-1/4}\,\lambda^{-1/2}\,e^{-t^2/(2\lambda^2)} \quad (15)$$

and, more specifically, to the "circular case" where $\lambda = 1$ for which the STFT admits a Bargmann factorization [10]

$$F_x^h(z) = \mathcal{F}_x^h(z)\,e^{-|z|^2/4}, \quad (16)$$

where $\mathcal{F}_x^h(z)$ is an entire function of the complex variable $z = \omega + \jmath t$. This also allows for a computationnally less expensive evaluation of the centroïds—involving only two STFT's since $t.h(t)$ and $dh(t)/dt$ happen to be proportional—, as well as closed forms expressions for the reassigned spectrogram in some simple cases such as, e.g., the model (12). In particular, if we let

$$\alpha = -\frac{1}{2}\left(\frac{1}{T^2} - i\,a\right)t^2 \quad (17)$$

and $\beta = 0$, it can be shown [8] that

$$\lim_{T\to\infty}\hat{S}_x^h(t,\omega) = \frac{1}{2\pi}\delta(\omega - at). \quad (18)$$

This examplifies the fact that, for unimodular linear chirps, the reassigned spectrogram is perfectly localized along the instantaneous frequency line, whatever the slope, exactly as the Wigner distribution does.

Another interesting special case is given by the Gabor logons corresponding to (17) with $a = 0$ (i.e., no chirping term). Such waveforms have minimum uncertainty, and an explicit calculation leads to

$$S_x^h(t,\omega) = \sqrt{\pi}\,e^{-\frac{1}{2}(t^2+\omega^2)} \Rightarrow \Delta_{t\omega}(S_x^h) = 2, \quad (19)$$

which, in accordance with (7), does correspond to the Heisenberg TF limit for a spectrogram. If a similar behavior is observed for the Wigner distribution:

$$W_x(t,\omega) = 2\sqrt{\pi}\,e^{-(t^2+\omega^2)} \Rightarrow \Delta_{t\omega}(W_x) = 1, \quad (20)$$

in accordance this time with (6), a more surprising result is obtained for the corresponding reassigned spectrogram, since the explicit calculation ends up with:

$$\hat{S}_x^h(t,\omega) = 4\sqrt{\pi}\,e^{-2(t^2+\omega^2)} \Rightarrow \Delta_{t\omega}(\hat{S}_x^h) = \frac{1}{2}, \quad (21)$$

with the paradox that the TF spread seems to be divided by two as compared to the Heisenberg limit! In fact, there is no real paradox and Heisenberg is not defeated by this sharp localization because reassignment has to be understood as a whole, characterized not only by the reassigned distribution but also by the vector field attached to the spectrogram values that have been moved. As for the classical Fourier analysis where localization has not to be confused with resolution, obtaining a sharp peak for one single component does not necessarily means the possibility of separating two closely spaced components. More precisely, it has been established [11] that, in the Gaussian case considered here, the reassignement vector field $\mathbf{r}_x(t,\omega) = (\hat{t}_x(t,\omega) - t, \hat{\omega}_x(t,\omega) - \omega)^t$ satisfies

$$\mathbf{r}_x(t,\omega) = \frac{1}{2}\nabla\log S_x^h(t,\omega), \quad (22)$$

with the further consequence that, in the reassignment process, the values of the spectrogram are moved along trajectories that point towards its local maxima [11].

This clearly evidences that the ultimately squeezed reassigned distribution results from contributions within a basin of attraction whose extent is precisely that of the original spectrogram, known to be constrained by uncertainty. Whenever more than one component would be present in such a (spectrogram) Heisenberg cell, the corresponding basins of attraction would then be competing, with interference patterns [6] but no super-resolution.

3. SPECTROGRAM GEOMETRY

3.1. On extrema

The above remark about basins of attraction, defined as domains surrounding local maxima of the log-spectrogram, suggests to have a closer look at the way such maxima are distributed in the TF plane. This can also be viewed as a dual problem of the distribution of zeros of the STFT (and, hence, of the spectrogram) that are known to entirely characterize the transform. Indeed, the analytic function $\mathcal{F}_x^h(z)$ that enters the Bargmann representation (16) being an entire function of order at most 2, it admits (up to some possible multiple zero at the origin) a Weierstrass-Hadamard factorization that is based on its zeros z_n and given by [12]:

$$\mathcal{F}_x^h(z) = e^{Q(z)}\prod_n(1-\tilde{z}_n)\,\exp\left(\tilde{z}_n + \tilde{z}_n^2/2\right), \quad (23)$$

where $Q(z)$ is a quadratic polynomial and $\tilde{z}_n = z/z_n$.

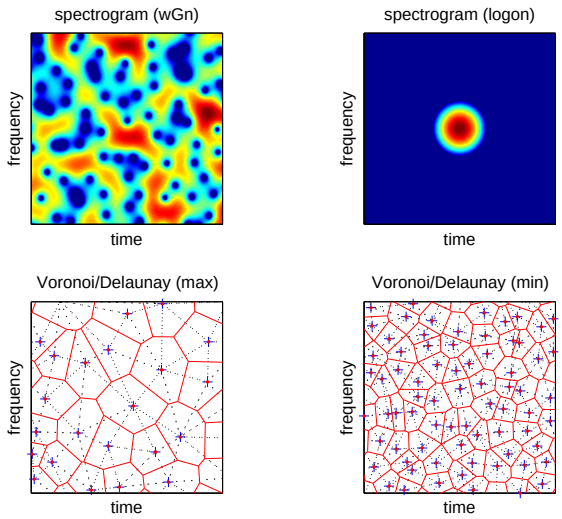

Fig. 2. *TF patches*. The spectrogram of one realization of white Gaussian noise (top left) is described in terms of Voronoi tessellations and Delaunay triangulations attached to the local maxima (bottom left) and local minima (bottom right). For a sake of comparison, the spectrogram of a Gabor logon with minimum uncertainty is also plotted (top right) with the same dynamic range (20 dB).

3.2. Voronoi and Delaunay

In order to address this question from an experimental point of view, we considered white Gaussian noise (wGn). The underlying idea is that, for any realization of wGn, a spectrogram should resemble a random distribution of "patches" whose shape and area should be controlled by the reproducing kernel of the analysis [2], i.e., the STFT of the analyzing window. A convenient way of identifying an approximation for the support of such patches is to construct the Voronoi diagram attached to the extrema. An example is given in Fig. 2. The simulation configuration corresponds to 256 data samples in the time domain, analyzed over 256 frequency bins. The length of the Gaussian window was chosen so as to match a "circular" geometry for the reproducing kernel and, in order to reduce border effects, only sub-squares of size 192×192 have been considered for further analysis.

3.3. A simplified model

Based on 100 independent realizations, a first result is that the average connectivity of maxima and minima with their nearest neighbourghs is, respectively, 5.90 and 5.98, i.e. almost 6. This suggests a simple model where maxima and minima would be located (on average) on a regular triangular lattice, with hexagonal Voronoi cells tiling the plane (see Fig. 3). Interestingly, such tiling of the plane is known to realize the maximum packing with circular patches. Some further interpretation is possible: if we associate a logon signal to each maximum, the interference pattern which results

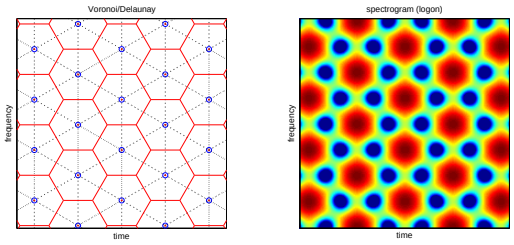

Fig. 3. *Average location of spectrogram extrema*. Model (left): the extrema (circled dots) are located on a triangular lattice, defining hexagonal Voronoi cells (red full lines), and the "circular" geometry of the analyzing window ensures that the distances between neighbouring extrema are all equal in the associated Delaunay triangulation (dotted black lines). Actual computation (right): when Gabor logons are centered on such a triangular grid, spectrogram maxima are located on this grid, while minima locations coincide with the nodes of the hexagonal lattice defined by the Voronoi tessellation.

from the interaction between any two such logons [2] yields two minima. We assume that the minima are located at edge nodes of the Voronoi cell. In this heuristic picture, maxima and minima are hence distributed over two entangled hexagonal lattices. A number of consequences can be drawn from this simple model and tested via numerical simulations. Some of them, involving 500 independent realizations, are reported in Fig. 4 with respect to *distance* between neigbouring extrema (as estimated from Delaunay triangulations), *number* of extrema and *area* of Voronoi cells measured in the normalized $(t/T, T\omega)$ plane, where T is the time scale for which (7) is an equality for Gaussian signals. In particular, if we let d_M and d_m be the distances between maxima and minima, N_M and N_m be the number of maxima and minima in a given domain and A_M and A_m be the area of Voronoi cells attached to maxima and minima, it can be derived from the model that:

$$d_M/d_m = \sqrt{3}; N_M/N_m = 1/3; A_M/A_m = 3. \qquad (24)$$

Albeit some significant dispersion is observed in the histograms of Fig. 4, mean experimental results are in a reasonable agreement with the theoretical predictions (24), thus supporting the choice of some randomized version of the model depicted in Fig. 3 for the locations and spreads of TF patches in a spectrogram of wGn. More can be said about the observed range of values for the considered attributes. We call "effective domain" of the minimum uncertainty logon (19) the circular domain which encompasses 95% of its energy. Its radius and area are equal to ~ 2.6 and 21.8 resp. (which is about 11 times larger than the Heisenberg spread given in (19)). From the typical size of the hexagonal cell of the model in Fig. 3, we deduce comparable values for both the radius $d_M/\sqrt{3} \sim 3$ and the area $2\pi/(3\sqrt{3})A_M \sim 21.8$.

Finally, as shown in Fig. 5, the distribution of areas closely resemble—when properly renormalized—that of the

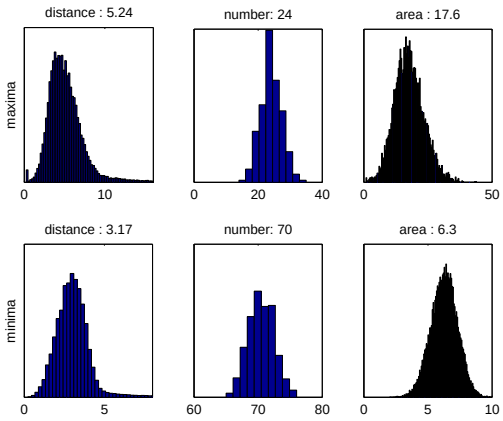

Fig. 4. *Geometrical attributes in Voronoi tessellations and Delaunay triangulations based on local extrema of wGn spectrograms within TF domains of size* 192×192. Histograms related to distance between extrema, number of extrema and area of Voronoi cells are plotted for both maxima (top row) and minima (bottom row), on the basis of 500 independent realizations. The empirical mean value of each distribution is indicated in each case.

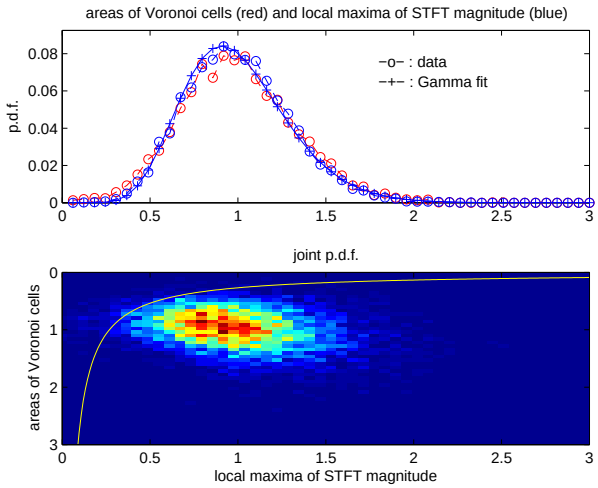

Fig. 5. *Distributions of local maxima and Voronoi cells areas in the case of wGn (simulation conditions as in Fig. 4).* When renormalized by their means, values of the local maxima of STFT magnitude and areas of the attached Voronoi cells are mildly correlated, with similar (Gamma-like) distributions (top: individual distributions; bottom: joint distribution). Uncertainty imposes furthermore that the product of those quantities is bounded from below: the yellow line stands for the boundary of the admissible domain.

corresponding local maxima of the STFT magnitude, with a correlation coefficient of 0.27. Albeit its explanation is still an open question, it can be further noticed that the latter can be well fitted by a Gamma distribution (with about 11 degrees of freedom). Starting from (7) and labelling $|F|_*$ the value of the local maximum of the STFT magnitude within a Voronoi cell of area A, it can be shown that $A.|F|_* \geq 3\sqrt{6}$, imposing some further, uncertainty-type, constraint on the coupling of those quantities. All those findings thus suggest that the average model of Fig. 3 could be refined according to

$$S_x^h(t,\omega) = \left| \sum_m \sum_n c_{mn}\, F_h^h(t-t_m, \omega-\omega_n) \right|^2, \qquad (25)$$

where the locations (t_m, ω_n) of the local maxima would be distributed on some suitably randomized version of the triangular grid and the weights c_{mn} would have their magnitude Gamma-distributed, with some partial correlation reflecting uncertainty constraints. This is currently under investigation and will be discussed elsewhere.

4. REFERENCES

[1] L. Cohen, *Time-Frequency Analysis*, Prentice Hall, 1995.

[2] P. Flandrin, *Time-Frequency / Time-Scale Analysis*, Academic Press, 1999.

[3] T.A.C.M. Claasen and W.F.G. Mecklenbräuker, "The Wigner distribution - A tool for time-frequency signal analysis. Part I: Continuous-time signals," *Philips J. Res.*, vol. 35, pp. 217–250, 1980.

[4] E. Schrödinger, "Zum Heisenbergschen Unschärfeprinzip," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, vol. 14, pp. 296–303, 1930, English translation available at http://arxiv.org/abs/quant-ph/9903100.

[5] D.F. Walls and G.J. Milburn, *Quantum Optics (2nd ed.)*, Springer, 2008.

[6] P. Flandrin, "Cross-terms and localization in time-frequency energy distributions," in *Time-Frequency Signal Analysis and Processing*, B. Boashash, Ed., chapter 4.2, pp. 94–101. Elsevier, 2003.

[7] K. Kodera, R. Gendrin, and C. de Villedary, "Analysis of time-varying signals with small *BT* values," *IEEE Trans. on Acoust., Speech and Signal Proc.*, vol. ASSP-26, no. 1, pp. 64–76, 1978.

[8] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. on Acoust., Speech and Signal Proc.*, vol. 43, no. 5, pp. 1068–1089, 1995.

[9] 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, Ed., chapter 5, pp. 179–203. CRC Press, Boca Raton, FL, 2003.

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

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

[12] 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.