PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
xxxxx, xxxxx, xxxx

Keywords:
xxxx, xxxx, xxxx

Author for correspondence:
Insert corresponding author name
e-mail: xxx@xxxx.xx.xx

Adaptive multimode signal reconstruction from time-frequency representations

Sylvain Meignen[1], Thomas Oberlin[2], Philippe Depalle[3], Patrick Flandrin[4], and Stephen McLaughlin[5]

[5]School of Engineering and Physical Sciences,Heriot-Watt University, Edinburgh Scotland, UK

[1] LJK, University of Grenoble, France

[2] IRIT and INP ENSEEIHT, University of Toulouse, France

[3] SPCL, McGill University, Montréal, Canada

[4] Laboratoire de Physique, ENS de Lyon, France

[5] Heriot-Watt University, Edinburgh, UK

This paper discusses adaptive reconstruction of the modes of multicomponent AM-FM signals by means of their time-frequency (TF) representation derived from their short time Fourier transform (STFT). The STFT of an AM-FM component or mode spreads the information relative to that mode in the TF plane around curves commonly called *ridges*. An alternative view, taken in this paper, is to consider a mode as a particular TF domain termed a *basin of attraction*. Here we discuss two new approaches to mode reconstruction, the first determines the ridge associated with a mode by considering location where the direction of the reassignment vector sharply changes; the technique used to determine the basin of attraction being directly derived from the method used for ridge extraction. A second uses the fact that the STFT of a signal is fully characterized by its zeros (and then the particular distribution of these zeros for Gaussian noise) to deduce an algorithm to compute the mode domains. For both techniques, mode reconstruction is then carried out by simply integrating the information inside these basins of attraction.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

In the signal processing community the meaning of frequency is well understood and over two centuries we have developed a series of mathematical tools to enable us to analyse signals in terms of the energy distribution in frequency. However, as with many obvious concepts, when we deal with signals which contain multiple components and these components are time-varying life becomes more difficult. The time-frequency (TF) analysis tools we have are often not appropriate or the results are problematic to interpret. In this paper we focus on the adaptive reconstruction of the modes of a multicomponent signal, consisting of AM-FM modes, by the use of the TF representation from the short time Fourier transform (STFT) of the signal.

The last 40 years have seen numerous TF methods proposed, (see, e.g., [1], [2] or [3] for surveys). In the methods for the analysis of AM-FM signals: the idealized scenario is that these signals correspond to a perfectly localized trajectory associated with the instantaneous frequency in the TF plane. Kodera, Gendrin and de Villedary pioneered an approach which modified the STFT [4]. They pointed out that spreading the STFT magnitude can be compensated for by taking into account the phase information usually discarded. Subsequently, Wigner-type distributions, tailored to guarantee localization of signals with specific FM laws were developed, though at the expense of cross-terms that were problematic in the multicomponent case. Auger and Flandrin (who coined the term *reassignment*) showed that the explicit use of the STFT phase can be efficiently replaced by a combination of STFTs with suitable windows [5]. Maes and Daubechies then developed *synchrosqueezing* [6] a special case of reassignment, with the additional advantage of enabling reconstruction. Such phase-based, data-driven methods have recently gained a renewed interest (see, e.g., the review paper [7]), and it is from here that the methods described in this paper begin.

The STFT of an AM-FM component or mode spreads the information relative to that mode in the TF plane around curves commonly called *ridges*. Conventionally the focus of signal reconstruction has been on dealing directly with these ridges. In this paper we develop an alternative view by considering a mode as a particular TF domain which we term a *basin of attraction*, (an early attempt in such a direction can be found in [8]). In this paper we focus on two approaches, the first approach determines the ridge associated with a mode by considering either the local maxima of the spectrogram in some predefined direction [9] or the zeros of the ridge points, in relation to *reassignment* techniques [5] [10]; the technique used to determine the basin of attraction is derived directly from the method for ridge extraction. The second exploits the fact that the STFT of a signal is fully characterized by its zeros [11] and then exploits the distribution of these zeros for Gaussian noise to deduce an algorithm which computes the basins of attraction. Mode reconstruction is then carried out by simply integrating the information inside these basins of attraction. Since, the zeros of STFT and the maxima of its modulus can both be used for mode reconstruction, the goal of this paper is to draw parallels and to highlight the differences between these two approaches and show their relevance in noisy situations and where the number of modes varies with time. One final remark to note is that the proposed approaches are expected to be beneficial for synchrosqueezing methods (see, e.g., [12] for recent advances), because determining basins is a pre-requisite for reconstruction.

The paper is structured as follows, first of all we introduce the basic analysis tools used for TF analysis such as the STFT and reassignment coupled with a discussion of the zeros of the STFT. Then the two different techniques which are the core of this paper and which compute the basins of attraction and mode domains are presented and discussed. We then present some numerical experiments to validate these methods when applied to a variety of synthetic and natural signals. Finally we draw some conclusions where we seek to highlight open issues worthy of further exploration.

2. Tools for time-frequency analysis

In this section, we briefly reprise the basics associated with TF analysis that are relevant to this paper, specifically: the STFT, ridges and their relationship to the modes of multicomponent signals and finally a brief description of the principles of reassignment. It then introduces by means of a brief discussion the notion of the zeros of the STFT.

(a) The short-time Fourier transform

Given a signal $f \in L^1(\mathbb{R})$, the space of real integrable functions, we define its Fourier transform by:

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(t)\, e^{-i\omega t}\, dt. \tag{2.1}$$

The *Short-Time Fourier Transform* (STFT) of signal f is then defined by

$$V_f^g(\omega, t) = \int_{\mathbb{R}} f(u) g(u-t) e^{-i\omega(u-t/2)}\, du, \tag{2.2}$$

where g is assumed to be a real-valued window with L^2 norm equal to 1. The spectrogram is then defined as $|V_f^g(\omega,t)|^2$. In the following, we will make extensive use of the unit energy Gaussian window defined as:

$$g(t) = \pi^{-1/4} e^{-t^2/2}. \tag{2.3}$$

(b) Ridges and multicomponent signals

A multicomponent signal is a superposition of AM/FM waves of the form:

$$f(t) = \sum_{k=1}^{K} f_k(t), \text{ with } f_k(t) = a_k(t) e^{i\phi_k(t)}, \tag{2.4}$$

for some finite K, where $a_k(t) > 0$ is a continuously differentiable function, ϕ_k is a two times continuously differentiable function satisfying $\phi_k'(t) > 0$ and $\phi_{k+1}'(t) > \phi_k'(t)$ for all t.

The general form for the STFT of a multicomponent signal admits the following first order approximation assuming $a_k'(t) \leq \varepsilon$ and $\phi_k''(t) \leq \varepsilon'$ [13,14]:

$$V_f^g(\omega, t) = \sum_{k=1}^{K} V_{f_k}^g(\omega, t) \approx \sum_{k=1}^{K} a_k(t) e^{i\phi_k(t)} \hat{g}(\omega - \phi_k'(t)). \tag{2.5}$$

We can further assume that the modes satisfy the following separation condition:

$$\phi_{k+1}'(t) - \phi_k'(t) \geq \Delta, \tag{2.6}$$

where Δ is called *separation parameter*. In this case, using equation (2.5) and assuming the essential frequency support of the window g is $[-\Delta, \Delta]$ the components occupy a distinct domain of the TF plane, allowing for their separation. In the following, f_k will be referred to as an AM-FM component. Going further $|V_f^g(\omega,t)| \approx \sum_{k=1}^{K} a_k(t) |\hat{g}(\omega - \phi_k'(t))|$, so that each mode f_k is associated with a TF *ridge* corresponding roughly to the $(t, \phi_k'(t))$ curve provided $|\hat{g}|$ attains its maximum at 0. The detection of ridges and their use in mode reconstruction has been pioneered in [13] and subsequently developed in a number of works (see, e.g., [14] or [15]).

(c) Reassignment basics

In this section, the principle underlying the *reassignment method* (RM) in the STFT context, which we will use later for ridge extraction, is introduced. The aim of RM is to compensate for the TF shifts induced by the 2D smoothing which defines the STFT. To do so, a meaningful TF location

to which the local energy given by the spectrogram is assigned, is first determined [5]. This corresponds to the *centroid* of the distribution, whose coordinates are defined by

$$\hat{t}_f(\omega,t) := \frac{t}{2} - \partial_\omega \arg V_f^g(\omega,t); \tag{2.7}$$

$$\hat{\omega}_f(\omega,t) := \frac{\omega}{2} + \partial_t \arg V_f^g(\omega,t), \tag{2.8}$$

at any point (ω,t) such that $V_f^g(\omega,t)\neq 0$. Both quantities, which *locally* define an instantaneous frequency and a group delay, enable perfect localization of linear chirps [5]. An efficient procedure computes them according to:

$$\hat{t}_f(\omega,t) = t + \Re\left\{\frac{V_f^{tg}(\omega,t)}{V_f^g(\omega,t)}\right\}; \tag{2.9}$$

$$\hat{\omega}_f(\omega,t) = \omega - \Im\left\{\frac{V_f^{g'}(\omega,t)}{V_f^g(\omega,t)}\right\}, \tag{2.10}$$

where tg stands for the function $tg(t)$ and $\Re\{Z\}$ (resp. $\Im\{Z\}$) is the real (resp. imaginary) part of the complex number Z. The RM is naturally associated with a so-called *reassignment vector* (RV) defined by:

$$RV(\omega,t) = \begin{pmatrix} \hat{t}_f(\omega,t) - t \\ \hat{\omega}_f(\omega,t) - \omega \end{pmatrix}. \tag{2.11}$$

Assuming the window g is the Gaussian window with unit variance defined in (2.3), then $V_f^{tg}(\eta,t) = -V_f^{g'}(\eta,t)$, and thus RV can be rewritten as:

$$\left(\Re\left\{\frac{V_f^{tg}(\omega,t)}{V_f^g(\omega,t)}\right\}, \Im\left\{\frac{V_f^{tg}(\omega,t)}{V_f^g(\omega,t)}\right\}\right)^t, \tag{2.12}$$

which can, in turn, be expressed in terms of the modulus of the STFT as follows [16]:

$$RV(\omega,t) = \nabla \log|V_f^g(\omega,t)|, \tag{2.13}$$

This last expression suggests a relationship between RV and local extrema of the amplitude of the STFT: $RV(\omega,t)$ is the null vector if and only if $|V_f^g(\omega,t)|$ admits a local extremum both along time and frequency directions. Moreover, (2.13) indicates that RVs tend to point towards local maxima that can be interpreted as attractors, whereas zeros act as repellers: this will be discussed and used in the following.

(d) On the zeros of the short time Fourier Transform

Rather than considering time and frequency independently, it can be interesting to consider them as the real and imaginary parts of a complex-valued variable, thus identifying the TF plane with the complex plane. Doing so by the introduction of $z=\omega+it$, allows a direct calculation to show that, when evaluated with the Gaussian window $g(t)$ defined in (2.3), the STFT (2.2) can be rewritten as:

$$V_f^g(\omega,t) = e^{-\frac{1}{4}|z|^2}\mathcal{V}_f(z), \tag{2.14}$$

where,

$$\mathcal{V}_f(z) = \int_{\mathbb{R}} A(z,s)f(s)ds, \tag{2.15}$$

with the kernel

$$A(z,s) := \pi^{-\frac{1}{4}} e^{-\frac{1}{2}s^2 - isz + \frac{1}{4}z^2}. \tag{2.16}$$

This corresponds to the Bargmann factorization of the STFT, with (2.15) the Bargmann transform of f [17], that happens to be an entire function of order 2. As a consequence, the Weierstrass-Hadamard theorem [18] guarantees that $\mathcal{V}_f(z)$ is completely characterized by the distribution of

its zeros, according to a factorization whose most general form reads

$$\mathcal{V}_f(z) = z^m e^{C_0+C_1 z+C_2 z^2} \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n}\right) e^{\frac{z}{z_n}+\frac{1}{2}\left(\frac{z}{z_n}\right)^2}, \tag{2.17}$$

where C_0, C_1 and C_2 are normalization, translation/rotation and squeezing factors respectively, and m is the possible multiplicity of a zero at the origin of the plane [19]. Since the zeros z_n of $\mathcal{V}_f(z)$ also correspond to the zeros of the STFT, the latter is fully characterized by its zeros.

3. Time-frequency segmentation for mode separation

In this section, we investigate two different techniques to compute the basin of attraction associated with one mode and then base our mode reconstruction on this approach. The first technique to compute the basins of attraction involves the use of the direction of RV to define the ridges and then the properties of the RV in the vicinity of the detected ridges, while the second is based on a study of the zeros of the STFT.

(a) Determination of basins of attraction based on ridges and reassignment vectors

There exist many different ways of computing the ridges associated with the TF representation given by the STFT [14,15]. The ridge detector proposed here is based on the properties of RV, in our context *ridge points* (RPs) are associated with a sharp variation in the direction of the RV. These can be computed by considering the projection of the RV onto a predefined direction associated with the angle $\frac{\pi}{2}+\gamma$ [9] and then the zeros of the projected vector, i.e.:

$$\Im\left\{\frac{V_f^{tg}(\omega,t)}{V_f^{g}(\omega,t)} e^{-i\gamma}\right\} = \langle RV(\omega,t), v_{\frac{\pi}{2}+\gamma}\rangle = 0, \tag{3.1}$$

where v_λ is the unit vector in the direction λ. The RPs thus correspond to points where the inner product of the RV with unit vector in the direction $\frac{\pi}{2}+\gamma$ changes sign.

Because $RV = \nabla \log |V_f^g|$ (see 2.13), $\langle RV(\omega,t), v_{\frac{\pi}{2}+\gamma}\rangle$ can be viewed as the directional gradient in the direction $\frac{\pi}{2}+\gamma$. Since in (3.1) the direction γ is fixed a priori, the method is not adaptive. To improve the adaptability of the method a variant was proposed in [10] which consisted of modifying the definition of RPs as follows:

$$\langle RV(\omega,t), v_{\theta(\omega,t) \bmod \pi}\rangle = 0. \tag{3.2}$$

with $RV(\omega,t) = r(\omega,t)e^{i\theta(\omega,t)}$ and where $\beta \mod \pi$ belongs to $[0,\pi[$. It is worth noting here that when RV belongs to $[0,\pi[$ (resp. $]-\pi,0]$), $\langle RV(\omega,t), v_{\theta(\omega,t) \bmod \pi}\rangle$ equals 1 (resp. -1).

The underlying rationale for this construction is that on each side of a ridge, the RV points in opposite directions. The behaviour of RV in the vicinity of a ridge associated with TF representation of Figure 1 A is displayed on Figure 1 C. So from one perspective, the ridge can be viewed as an attractor for the reassignment vector field. The RPs defined in (3.2) are not only associated with ridges but form more general structures in the TF plane which we call *contours*. These contours do not branch, but terminate on the borders of the TF axes, or form closed loops as discussed in [9]. In addition, the phase along a contour varies smoothly until it passes through a singularity in the zeros of the STFT. The method then segments the contours whenever they cross the zeros of the STFT. This raises the question, why do contours necessarily pass through zeros of the STFT? The RV as stated in expression (2.13) is oriented, in the vicinity of the zeros of the spectrogram of a bat signal (displayed in Figure 1 A), as shown in Figure 1 B. Note that this behaviour for RV in the vicinity of the zeros of the spectrogram is independent of the the signal studied. Consequently, the zeros can be viewed as reppellers for the reassignment vector field. Then, due to the "mod π" term in the definition (3.2), $\langle RV(\omega,t), v_{\theta(\omega,t) \bmod \pi}\rangle$ is positive above a zero and negative otherwise. Thus, the contour is horizontal in the vicinity of these points. Note

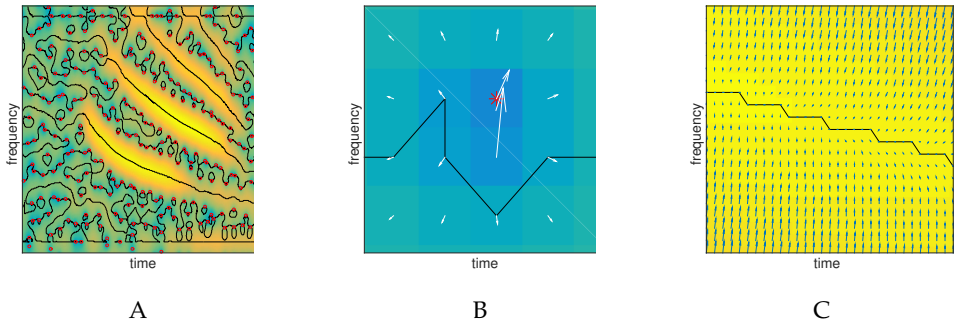

Figure 1. A: Signal spectrogram of a bat signal with the zeros and the ridges superimposed; B: RV close to a zero of the spectrogram (in red), white arrows represent the RV and the nearby contour is plotted in black; C: RV close to a ridge of the spectrogram, blue arrows represent the RV and the ridge is plotted in black.

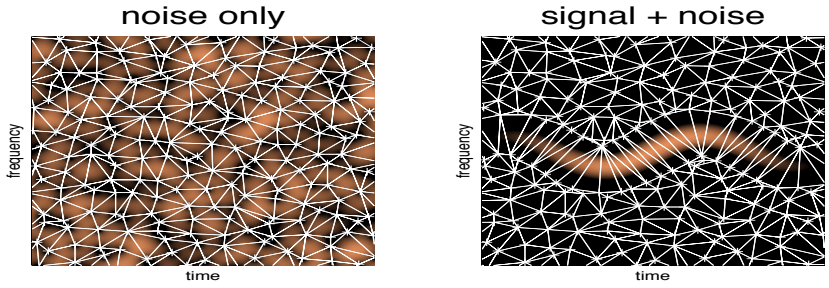

Figure 2. left: Delaunay triangulation based on the zeros of the spectrogram for a noise signal; right: Delaunay triangulation based on the zeros of the spectrogram of a mode superimposed to the noise of the left diagram.

that, due to TF discretization, zeros of the spectrogram cannot be exactly determined, therefore a contour actually passes in the vicinity of the discrete minimum but not necessarily onto it.

Another interesting aspect of the proposed method lies in the fact that the contour does not necessarily last for the whole time span of the signal: if the energy becomes locally too small the ridge ends on the nearby zeros of the STFT. This last point is very important since it indicates that a mode dies when it connects with a zero. Usually, the presence or otherwise of a mode is assessed in terms of the statistics of the spectrogram of the noise. The basic idea being that in order to qualify as a mode the points should not obey this statistic, and an ad hoc machinery is required to connect such points [20]. In our framework, it is not necessary to perform any statistical analysis to characterize the birth and death of a mode since our study is based on the locations of the zeros. An illustration of this phenomenon is given in Figure 1 A, where the contours and the zeros are superimposed on the spectrogram of a noisy bat signal.

Finally, to build the basin of attraction knowing the location of a contour, we consider the most frequent contour in the vicinity of $(\hat{t}_f(\omega,t),\hat{\omega}_f(\omega,t))$. Each point (ω,t) is then given the index of the ridge it is attached to, and the set of points attached to ridge i is denoted by $\mathcal{B}_i$. Note here that the notion of basin of attraction is one way to recover the information associated with one mode and there exist several other ways of gathering this information as will be discussed later.

(b) Determination of mode domains based on Delaunay triangulation upon zeros

The second approach to determining a TF domain attached to a given mode arises from noting that the zeros of the STFT completely characterize it. Consequently it is natural to consider these as a 2D point process in the TF plane, with properties which are associated with the nature of the analyzed signal. So initially we consider a simplified, geometrical description of the TF structure of a signal by considering the diagrams connecting STFT zeros, i.e. the so-called *stellar representation* used in Quantum Mechanics [19,21] (see also [22] for a related TF perspective). The simplest approach is to use Delaunay triangulation [23]. When these diagrams are considered for white Gaussian noise using the Delaunay triangulation approach, a homogeneous 2D random field defined by the STFT zeros results and the distribution of the zeros is itself homogeneous over the plane. It is expected that this will be no longer the case for a signal with a coherent TF structure, e.g. frequency modulation superimposed. Figure 2 illustrates that this is exactly what happens: when an AM-FM chirp is added to the noise of the left diagram, the noise-only regions remain unaffected whereas the signal domain is characterized not only by large spectrogram values but also by Delaunay triangles with longer edges than in noise-only regions. Note that the addition of coherent signal structure to incoherent noise does not change the total number of zeros in the plane, but simply modifies their distribution so that they tend to align along the border of mode domains that tend to identify to the previously defined basins of attraction.

Theoretical considerations and the evidence displayed in Figure 2 suggest that signal domains could be identified by considering Delaunay triangles that depart from the expected behaviour associated with noise. In [11] it was shown that the distribution of edge lengths of Delaunay triangles constructed using STFT zeros for the case of white Gaussian noise is essentially bounded above by a maximum value $L_{\max} \sim 2.2$ (in the system of normalized units corresponding to the window (2.3)), with a very low probability of exceeding 2 (referring as $|e_{mn}|$ the distance between any two zeros z_m and z_n in a Delaunay triangle, a numerical evaluation evidenced that $\mathbb{P}\{|e_{mn}| > 2\} \approx 10^{-3}$). Selecting Delaunay triangles on the basis of thresholding their maximum edge length is therefore a simple way of identifying elementary local domains whose concatenation defines global mode domains $\mathcal{B}_i$ via supports delineated by zeros.

(c) Mode reconstruction

Whatever the method (ridges or zeros), having computed the different basins $\mathcal{B}_i$ corresponding to the different modes of the signal, the reconstruction of the mode associated with basin i is then performed by means of the classical inversion formula:

$$f_i(t) = \int\int_{\mathbb{R}^2} V_f(\omega, t) \mathbb{1}_{\mathcal{B}_i}(\omega, t) g(\tau - t) e^{i\omega(\tau - t/2)} dt d\omega / 2\pi, \tag{3.3}$$

where $\mathbb{1}_X$ is the indicator function of X.

(d) Comparison of the ridge and zeros based method for the computation of basins of attraction

We now compare the techniques based on the computation of the basins of attraction using ridge location and RV direction to that based on the mode domains resulting from the Delaunay triangulation of the zeros of the spectrogram. To illustrate the different behaviours we first consider a synthetic noisy three mode signal, each mode having sinusoidal phase. We display in Figure 3-A the spectrogram of the signal along with the ridges and basins of attraction corresponding to each mode on Figure 3-B and finally the mode domains based on Delaunay triangulation for the same noisy signal on Figure 3-C. We notice that as expected the TF domains associated with the three modes are correctly estimated allowing a good separation of the modes even when the noise level is relatively high (10 dB in this example).

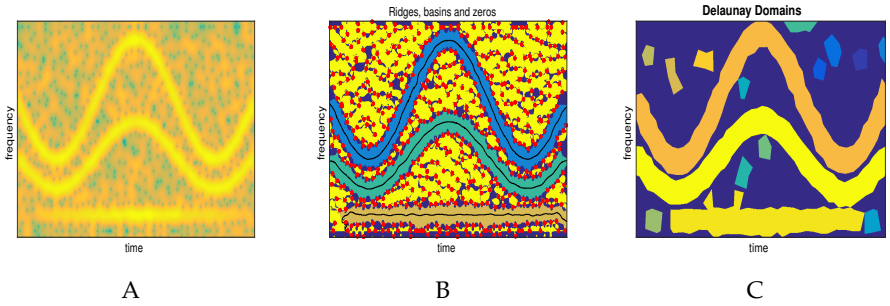

Figure 3. A: Spectrogram of synthetic three mode signal with additive Gaussian noise (SNR = 10 dB); B: Ridges and basins of attraction computed as explained in section 3(a); C: Mode domains computed using Delaunay triangulation as explained in section 3(b)

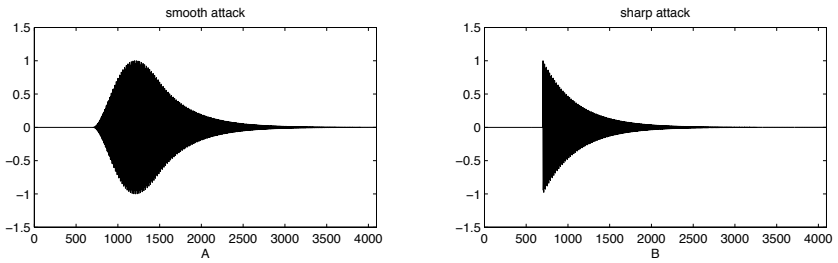

Figure 4. Damped tone with a smooth attack (raised cosine, left column) and a sharp attack (step function, right column).

4. Numerical validation

In this section, we investigate the properties of the proposed techniques for mode identification and reconstruction for signal denoising as well as illustrating a sound processing application [24]. We have selected two types of signals as examples for their intrinsic TF properties that fit the AM/FM multi-component model well, cf. Eq. (2.4) but also for their ability to cover a large class of musical sounds [25] (and more generally the class of signals produced by vibrating structures [26]). These are: the damped sinusoid (with a smooth or a sharp attack) related to percussive signals, i.e. the impulse response of a resonant structure, and an excerpt of a cello sound, a typical case of a sustained harmonic signal [27].

(a) Denoising of a damped tone

We first study the denoising performance of the proposed algorithms based on determining the basins of attraction and mode domains for reconstruction, on a damped tone, a simple and widely used model in audio. [28]. Two cases are considered, with either a smooth attack (raised cosine) or a sharp one (step function), see respectively Figures 4 A and 4 B.

In the first case, we display the spectrogram of such a signal contaminated by additive white Gaussian noise on Figure 5 A, the basins of attraction with respect to the energy of the STFT computed on their associated ridges on Figure 5 B and the Delaunay domains on Figure 5 C. Finally, Figure 5-D presents the denoising performance for both methods. The output SNR is computed by comparing the original signal to the mode reconstructed using the 3 most energetic basins of attraction for the method based on RV or Delaunay triangulation. For the Delaunay triangulation method, the choice of considering the 3 most energetic basins of attraction is based on the numerical observation: most of the energy is contained in the first three basins of attraction. In this example, both methods behave in a very similar fashion.

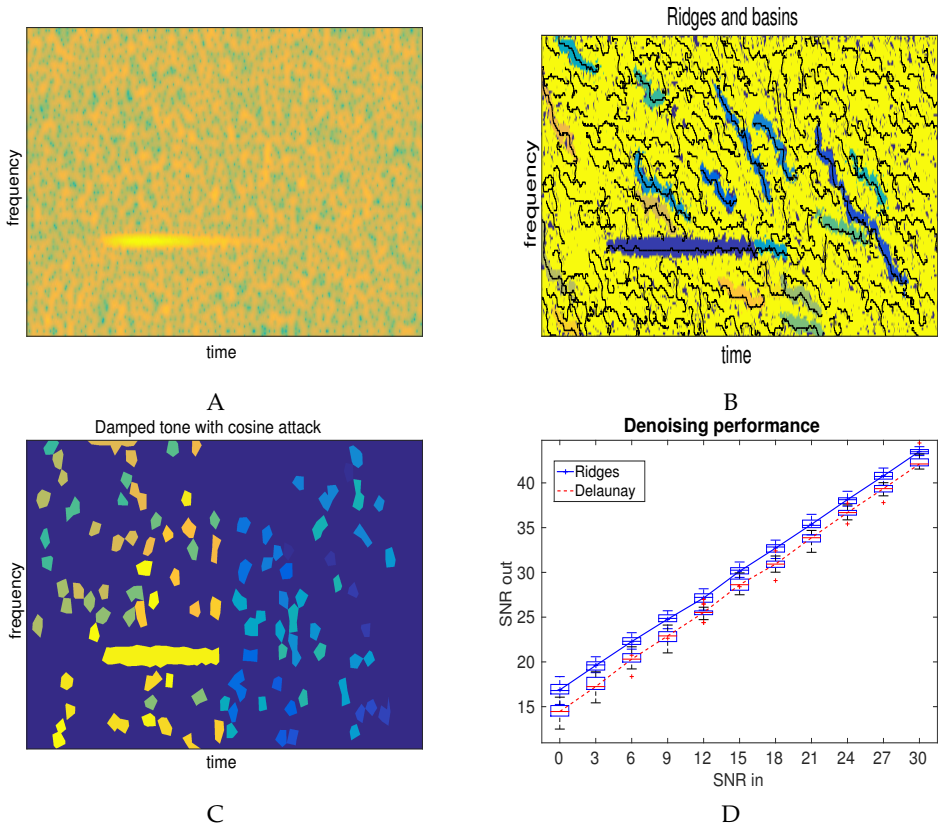

Figure 5. A: Spectrogram of a noisy damped tone with a smooth attack (SNR = 0 dB); B: Main ridges and basins of attraction derived from the RV; C: Mode domains derived from Delaunay triangulation; D: Denoising performance of the mode reconstruction.

A similar analysis is provided for the case of the sharp attack, see Figure 6. The damped tone with a sharp attack exhibits a TF representation made of a horizontal and a vertical part, the latter corresponding to the attack. What is interesting about this signal is that it no longer satisfies (2.4) because the amplitude of the mode is no longer differentiable. The behaviour of the two methods on this type of signal are significantly different. While the method based on the Delaunay triangulation manages to capture the horizontal and the vertical parts of the mode, the method based on RV never acquires the attack (even at low noise levels). This results in an output SNR which is much worse for the latter compared with the former method. Note that the vertical TF structure could be detected using the method based on RV by changing $\theta(\omega, t) \bmod \pi$ in (3.2) into $\theta(\omega, t) + \frac{\pi}{2} \bmod \pi - \frac{\pi}{2}$, but then two different analyses are required to properly analyze such a signal. An alternative method, connected with the RV, which will be investigated in the near future, would be based on the analysis on bidimensional ridge detection [29].

(b) Analysis of a cello sound

We now apply the techniques proposed in this paper to the problem of identification and extraction of the harmonic structure of the first part of a cello note, a G5 (a frequency of $776Hz$). The temporal structure of the example (sampled at 11025 Hz and lasting for 0.7 seconds) is relatively complex and shows a fast attack followed by some AM-FM modulations (cf. Figure 7-A). Its spectrograms is easier to decipher as it clearly exhibits the first six partials that lie within the available frequency range $[0, 55012.5]$ Hz (Figure 8-A). Note the vibrato effect, which appears as a

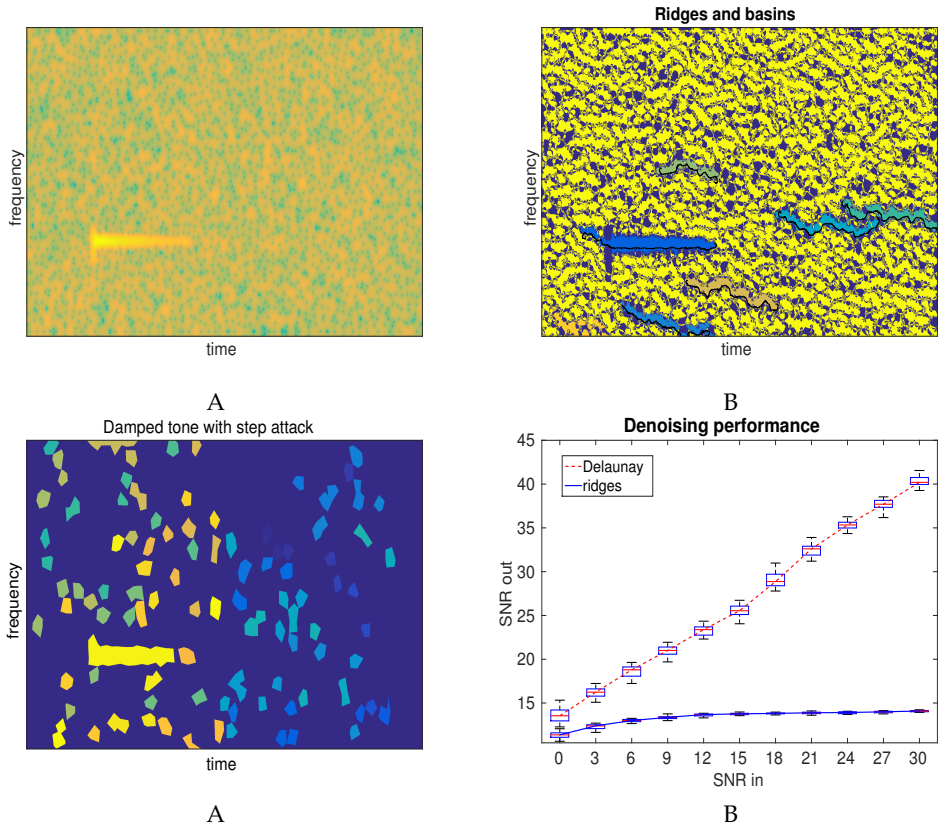

Figure 6. A: Spectrogram of a noisy damped tone with a sharp attack (SNR = 0 dB); B: Main ridges and basins of attraction derived from the RV; C: Mode domains derived from Delaunay triangulation; D: Denoising performance of the mode reconstruction methods

pseudo-sinusoidal modulation of the fundamental frequency and results in a frequency deviation that increases proportionally to the harmonic rank. Also note the presence of noise during the note attack around 0.05s; this is produced by the bow rubbing the string while the oscillatory motion takes place [27].

For the RV-based analysis (Figure 8-B), we verify once again that the basins of attraction are easily determined for high amplitude modes; this is indeed the case for the first (i.e., the fundamental) and the fourth. For the other modes, spurious zeros in the spectrogram might cause the splitting of the basin associated with a single mode into several smaller basins. Note that while the algorithm detects the modes, reconstructing them based on their basins of attraction would require some post-processing steps in order to group them in some relevant way.

In the case of the mode domain extraction based on the Delaunay triangulation, each of the six harmonic partials are clearly identified by corresponding domains that track frequency modulations well (Figure 8-C). The highest amplitude partial is entirely represented by only one domain, while each of the others is represented by a few sub-domains, usually due to a missing triangle that breaks the continuity of the whole. As in the case of the RV-based method, a post-processing step would help merging sub-domains to reconstitute the domain that represents a whole partial. In addition, the Delaunay-based method confirms its ability to detect and render transients. Indeed, note the vertical structure centred around the attack time (0.05s, Figure 8-C). Finally, an experiment was performed to reconstruct the harmonic and attack structure of the cello sound from selected Delaunay domains applied as masks through reconstruction technique (3.3). One can see the temporal representation of the reconstructed signal in Figure 7-B which looks

very similar to the original excerpt (cf. Figure 7-A). An auditory comparison confirms how close the two signals are.

In terms of post-processing, for both of the RV and Delaunay-based methods, notice that when the modes are harmonically related, one can benefit from the fact that the instantaneous frequencies of harmonic components are multiples of the fundamental frequency to gather together basins of attraction/mode domains associated with a given mode. Notice also that this pertains to the problem of reconstructing trajectories of partials in the context of additive synthesis [30]. This is a topic which we will seek to address in the near future. A fruitful application for this harmonic identification/extraction is the separation of the noisy part from the pseudo-periodic part of any given sound. This is a key step in preserving the perceptual naturalness and in achieving high quality results in advanced audio processing techniques such as time-scaling, transposition or sound morphing [31]. Indeed, one has to process differently the pseudo-deterministic part from the noisy part (e.g., for time-scaling and transposing sounds, the noisy attack part has to remain unchanged).

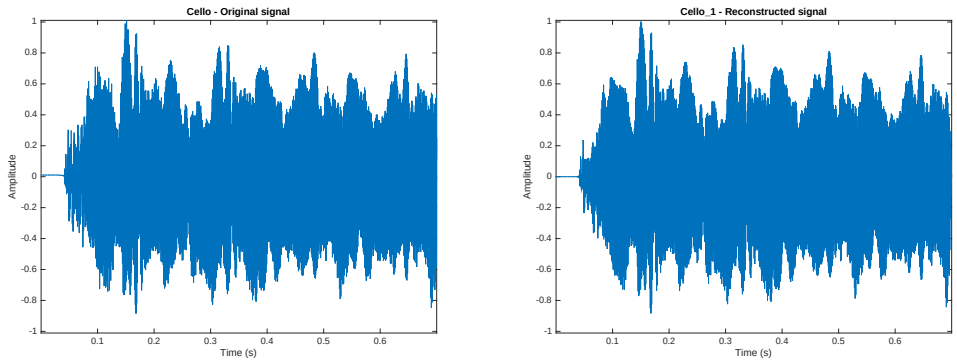

Figure 7. A: Cello sound - original signal; B: Cello Sound - reconstructed signal after mode domain extraction derived from Delaunay triangulation.

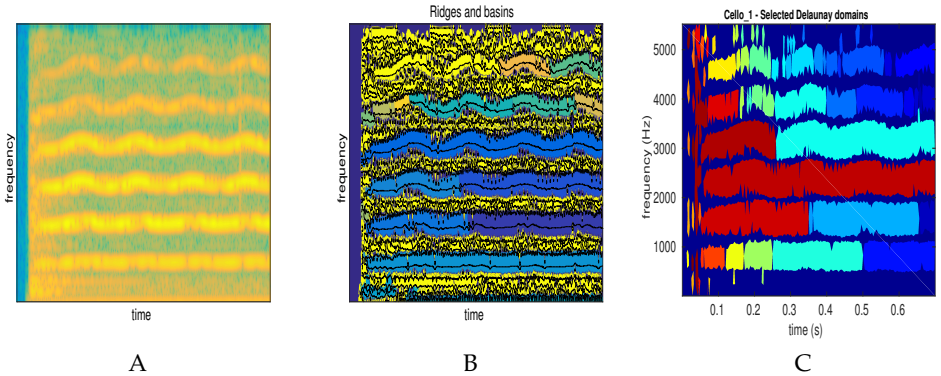

Figure 8. A: Spectrogram of a cello sound; B: Main ridges and basins of attraction derived from the RV; C: Mode domains derived from Delaunay triangulation.

5. Conclusion

In this paper we have presented two different STFT-based mode decompositions: the first estimates the *ridges*, then computes the basins associated with the modes by making use of the reassignment vector. The second is more novel, since it only uses the location of the STFT's zeros to build these domains. Both methods provide a segmentation of the TF-plane into meaningful components, and a reconstruction of these modes in a fully adaptive way. The experiments presented here show that both methods can decompose multicomponent signals with high accuracy, even when the signals are contaminated by noise. The comparisons also show that the ridge-based reconstruction can be more precise, whereas the Delaunay-based segmentation allows for more complex modes (e.g. containing sharp attacks, or impulses). To conclude, we note that such TF adaptive decompositions can be easily applied to time-scale representations (e.g., continuous wavelet transform), which is more appropriate for some particular modulations and inter-mode separation. Similarly, it should be possible to adapt such decompositions to non-harmonic signals. This will be the subject of further study, which will also focus on real-life applications e.g. in automatic music signal analysis.

Funding statement. The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under grant ANR-13-BS03-0002-01 (ASTRES) and of the EPSRC via grant EP/J015180/1.

Conflict of interests. We have no competing interests

Authors' contributions. SML, SM and PF conceived of writing the manuscript. SM and TO developed the ridge based method and carried out associated experiments. PF initiated the Delaunay based method, and PD and PF developed the numerical implementation and carried out the associated experiments. All authors developed and contributed to the writing and editing of the manuscript.

References

1. Flandrin P. 1999 *Time-Frequency/Time-Scale Analysis*.
   Academic Press.
2. Gröchenig K. 2001 *Foundations of Time-Frequency Analysis*.
   Springer.
3. Boashash B. 2003 *(ed.) Time-Frequency Signal Analysis and Processing — A Comprehensive Reference*.
   Elsevier.
4. Kodera K, De Villedary C, Gendrin R. 1976 A new method for the numerical analysis of non-stationary signals.
   Phys. Earth Plan. Inter. **12**, 142–150.
5. Auger F, Flandrin P. 1995 Improving the readability of time-frequency and time-scale representations by the reassignment method.
   IEEE Transactions on Signal Processing **43**, 1068–1089.
6. Daubechies I, Maes S. 1996 A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models.
   In *Wavelets in Medicine and Biology* (ed. A Aldroubi, M Unser), pp. 527–546. CRC press.
7. Auger F, Flandrin P, Lin Y, McLaughlin S, Meignen S, Oberlin T, Wu H. 2013 Time-frequency reassignment and synchrosqueezing.
   IEEE Signal Processing Magazine **30**, 32–41.
8. Chassande-Mottin E, Auger F, Flandrin P, Daubechies I. 1997 Partition du plan temps-fréquence et réallocation.
   In *Actes du Colloque GRETSI-97*, pp. 1447–1450. Grenoble, F.
9. Lim Y, Shinn-Cunningham B, Gardner T. 2012 Sparse contour representations of sound.
   IEEE Signal Processing Letters **19**, 684–687.
10. Meignen S, Gardner T, Oberlin T. 2015 Time-frequency ridge analysis based on reassignment vector.
    In *Proceedings of the 23st European Signal Processing Conference (EUSIPCO-15)*. (to appear).
11. Flandrin P. 2015 Time-frequency filtering based on spectrogram zeros.
    submitted, preprint at https://hal-ens-lyon.archives-ouvertes.fr/ensl-01121522 .

12. Daubechies I, Lu J, Wu H. 2011 Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool.
Applied and Computational Harmonic Analysis **30**, 243–261.
13. Delprat N, Escudié B, Guillemain P, Kronland-Martinet R, Tchamitchian P, Torrésani B. 1992 Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies.
IEEE Transactions on Information Theory **38**, 644–664.
14. Carmona R, Hwang W, Torresani B. 1997 Characterization of signals by the ridges of their wavelet transforms.
IEEE Transactions on Signal Processing **45**, 2586–2590.
15. Newland D. 1999 Ridge and phase identification in the frequency analysis of transient signals by harmonic wavelets.
Journal of Vibration and Acoustics **121**, 149–155.
16. Chassande-Mottin E, Daubechies I, Auger F, Flandrin P. 1997 Differential reassignment.
IEEE Signal Processing Letters **4**, 293–294.
17. Bargmann V. 1961 On a Hilbert space of analytic functions and an associated integral transform Part I.
Communications on Pure and Applied Mathematics **14**, 187–214.
18. Boas R. 1954 *Entire Functions*.
Academic Press, New York.
19. Korsch H, Muller C, Wiescher H. 1997 On the zeros of the Husimi distribution.
Journal of Physics A: Mathematics and General **30**, 677–684.
20. Huillery J, Millioz F, Martin N. 2014 Gaussian noise time-varying power spectrum estimation with minimal statistics.
IEEE Transactions on Signal Processing **62**, 5892–5906.
21. Lebœuf P, Voros A. 1990 Chaos-revealing multiplicative representation of quantum eigenstates.
Journal of Physics A: Mathematics and General **23**, 1765–1774.
22. Gardner TJ, Magnasco MO. 2006 Sparse time-frequency representations.
Proceedings of the National Academy of Sciences **103**, 6094–6099.
23. Okabe A, Boots B, Sugihara K, Chiu S. 2000 *Spatial Tessellations — Concepts and Applications of Voronoi Diagrams (2nd ed.)*.
John Wiley.
24. Depalle P, Kronland-Martinet R, Torrésani B. 2007 Time-frequency multipliers for sound synthesis.
In *Proceedings of the Wavelet XII Conference, SPIE Annual Symposium*, pp. 670118–1–670118–15. San Diego, CA.
25. Quatieri T, McAulay R. 2002 Audio signal processing based on sinusoidal analysis/synthesis.
In *Applications of Digital Signal Processing to Audio and Acoustics* (ed. M Kahrs, K Brandenburg), volume 437 of *The International Series in Engineering and Computer Science*, pp. 343–416. Springer.
26. Den Hartog J. 1985 *Mechanical Vibrations (4th ed.)*.
New York: Dover Publications.
27. Fletcher NH, Rossing TD. 1998 *The Physics of Musical Instruments (2nd ed.)*.
Springer Verlag.
28. Rodet X. 1984 Time-domain formant-wave-function synthesis.
Computer Music Journal **8**, pp. 9–14.
29. lindeberg T. 1998 Edge detection and ridge detection with automatic scale selection.
Internatioanl Journal of Computer Vision **30**, 117–156.
30. Kereliuk C, Depalle P. 2008 Improved hidden markov model partial tracking through time-frequency analysis.
In *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, pp. 111–116. Espoo, FI.
31. Arfib D, Keiler F, Zoelzer U, Verfaille V, Bonada J. 2011 Time-frequency processing.
In *DAFX: Digital Audio Effects, 2nd Edition* (ed. U Zoelzer), pp. 219–278. John Wiley & Sons, Ltd, Chichester, UK.