Wolf population counting by spectrogram image processing [1]

B. Dugnol, C. Fernández and G. Galiano [2]

Dpto. de Matemáticas, Universidad de Oviedo, c/ Calvo Sotelo, 33007-Oviedo Spain

Abstract

We investigate the use of image processing techniques based on partial differential equations applied to the image produced by time-frequency representations of one-dimensional signals, such as the spectrogram. Specifically, we use the PDE model introduced by Álvarez, Lions and Morel for noise smoothing and edge enhancement, which we show to be stable under signal and window perturbations in the spectrogram image. We demonstrate by numerical examples that the corresponding numerical algorithm applied on the spectrogram of a noisy signal reduces the noise and produce an enhancement of the instantaneous frequency lines, allowing to track this lines more accurately than with the original spectrogram. We apply this technique both for synthetic signals and for wolves chorus field recorded signals, which was the original motivation of this work.

Key words: Spectrogram, time-frequency distribution, noise, partial differential equation, instantaneous frequency, image processing, population counting.

1 Introduction

Wolf is a protected specie in many countries around the world. Due to their predator character and to their proximity to human settlements, wolves often kill cattle interfering in this way in farmers' economy. To smooth this interference, authorities reimburse the cost of these lost to farmers. Counting the

[1] All the authors are partially supported by Project PC0448, Gobierno del Principado de Asturias, Spain. The third author is partially supported by the Spanish DGI Project MTM2004-05417.

[2] Corresponding author. Phone:+34 985182299 Fax: +34 985103354

[3] dugnol@ieee.org, carlos@orion.ciencias.uniovi.es, galiano@uniovi.es

population of wolves inhabiting a region is, therefore, not only a question of biological interest but also of economic interest, since authorities are willing to estimate the budget devoted to costs produced by wolf protection, see for instance [1]. However, estimating the population of wild species is not an easy task. In particular, for mammals, few and not very precise techniques are used, mainly based on the recuperation of field traces, such as steps, excrements and so on. In this article we propose what it seems to be a new technique to estimate the population of species which fulfill two conditions: they live in groups, for instance, packs of wolves, and they emit some characteristic sounds, howls and barks, for wolves. This technique consists of three steps. The first step is the spatial localization of wolves packs and the recording of their chorus. Despite the quality of the recording devices, these recordings are affected for a variety of undesirable signals, that we shall call *noise*, which range from low amplitude broad spectrum long duration signals, like wind, to more localized signals in time, like cattle bells, or more localized in spectrum, like car engines. Clearly, the addition of all these signals generates an unstructured noise in the background of the wolves chorus.

Reducing noise in the recorded signal is the second step and the main issue of this article. Before dealing with this subject let us explain the third step, which consists on identifying how many different voices are emitting in a given recording, task that can be seen as a simplified version of speech recognition, and that we shall approach by instantaneous frequency estimation using time-frequency analysis. Time-frequency analysis has proven to be an effective tool for analyzing the behavior of non-stationary signals such as speech, music and biological and geophysical signals, among others. By displaying a signal over a joint time-frequency plane using a time-frequency representation it is possible to reveal certain structures that are not apparent neither in the time domain nor in the frequency domain alone. The most popular analysis tool in many cases is the spectrogram, which has been extensively used in speech analysis, see for instance [2,3].

A time-frequency representation, like the spectrogram, which is essentially an energy distribution in the time-frequency plane, ideally must be localized along the instantaneous frequency of the signal. Instantaneous frequency is a notion which generalizes the usual notion of frequency for signals with frequencies changing in time. For instance, a cosine modulation $x(t) = a\cos(\omega_0 t + \phi_0)$ has a frequency ω_0 that is the derivative of the phase $\phi(t) = \omega_0 t + \phi_0$. To generalize this notion, real signals x are written as an amplitude, a, modulated with a time varying phase, ϕ:

$$x(t) = a(t)\cos\phi(t),$$

with $a(t) \geq 0$. Then, the *instantaneous frequency* is defined as a positive derivative of the phase: $\omega(t) = \phi'(t) \geq 0$. In our context, and for a clean recording, instantaneous frequencies represent the different voices (up to harmonics of a given tone) contained in a chorus of wolves and thus, their esti-

mation correspond to an estimation of the number of wolves singing in that chorus.

Time-frequency representations may be divided into parametric and nonparametric methods. Among nonparametric approaches, the Wigner-Ville distribution has received much attention due to its excellent concentration and many other desirable mathematical properties, see [3]. However, it is well known that the Wigner-Ville distribution presents high amplitude cross-terms for multi-component signals which makes its interpretation difficult. Various distributions have been introduced to reduce cross-term while preserving a high time-frequency resolution, but none of them give satisfactory results in all possible applications, see [3–6,8]. Parametric approaches assume a priori knowledge about the signal. This is the case in a wide range of scenarios where the signal to be analyzed can be assumed to follow some model and then the problem becomes that of estimating the parameters of that model, see for instance [7,8].

As mentioned above, the main concern of this article is the noise reduction of the signal for latter estimation of instantaneous frequencies. Since the noise is unstructured, it is difficult to obtain a good general denoising algorithm working only in the time domain. In addition, being the instantaneous frequency estimation our objective, it seems reasonable to work directly with a time-frequency energy representation and to apply our denoising algorithm to the image produced by that representation, which should be robust with respect to noise perturbation and to the multi-component character of even an ideal clean signal containing only the wolves chorus. It seems then clear that Wigner-Ville type distributions are not adequate for our purpose. Parametric representations are not to be rejected, specially those related to chirplet transforms which, in fact, is a line of current research in our work. However, we prefer to present here our results for the most commonly used representation, the spectrogram, which presents good properties of robustness against noise corruption and multi-component character of signals as well as having a very intuitive interpretation and a low computational cost.

Considering the spectrogram of a signal corrupted with noise as an image, we propose in this article the use of an image processing technique for edge enhancement and noise reduction based on a regularization of the mean curvature motion equation, as introduced in [9]. There exist a variety of PDE-based models for smoothing and enhancing images that could be used instead of that introduced in [9]. We refer the readers to the books [10,11] for further insight into the problem. In few words, the parabolic partial differential equation introduced in [9], takes, in our application, the spectrogram, S, of a signal as initial data and transform it in the following way: in points (t, ω) in which the mean of $|\nabla S|$ in a neighborhood of (t, ω) is large (edges of instantaneous frequency strips), diffusion is almost inhibited and edges are therefore preserved.

If this mean is small, diffusion takes place according to two main situations. If $|\nabla S|$ is relatively large, diffusion on the orthogonal direction to the gradient is applied, resulting in an enhancement of the edge. When $|\nabla S|$ is small, isotropic diffusion takes place, making the neighborhood of the point more homogeneous.

Finally, let us mention that signal theory is a well known tool in other branches of animal behavior research such as the analysis of acoustic recognition (whale songs [12], dolphin whistles [13], penguin sounds [14]), or of space location, as in bats [15], for which our technique could also be applied.

2 An outline of the method

In this section we present a simple example aimed to introducing the main intuitive ideas underlying the technical aspects of our method. Therefore, rigorous mathematical arguments will be substituted by approximate calculations for simplicity in the exposition.

Let $x \in L^2(\mathbb{R})$ denote an audio signal and consider the Gabor's transform $\mathcal{G} : \mathbb{R}^2 \to \mathbb{C}$ given by

$$\mathcal{G}x(t,\omega) = \int_{\mathbb{R}} x(s)\varphi(s-t)e^{-i\omega s}ds, \tag{1}$$

corresponding to the real, symmetric and normalized *window* $\varphi : \mathbb{R} \to \mathbb{R}$. The energy density function or *spectrogram* of x corresponding to the window φ is given by

$$Sx(t,\omega) = |\mathcal{G}x(t,\omega)|^2. \tag{2}$$

Assume that the signal x is composed by the addition of two signals, $x = \lambda_1 x_1 + \lambda_2 x_2$, for positive λ_1 and λ_2. Let x_1 be a sinusoidal complex wave, representing a wolf's steady howl, $x_1(t) = \exp(i\omega_0 t)$, and x_2 be a white Gaussian noise. We have taken x_1 complex for simplicity in the computations, but it could be given by a real signal like $\cos(\omega_0 t)$. For a Gaussian window of variance σ, $\varphi(t) = (\pi\sigma^2)^{-1/4}\exp[-t^2/(2\sigma^2)]$, the Gabor's transform of x_1 is given by $\mathcal{G}x_1(\omega,t) = \hat{\varphi}(\omega-\omega_0)\exp[-it(\omega-\omega_0)]$, with $\hat{\varphi}(\omega) = (4\pi\sigma^2)^{1/4}\exp[-\sigma^2\omega^2/2]$, and then

$$Sx_1(t,\omega) \equiv Sx_1(\omega) = 2\sqrt{\pi}\sigma\exp[-\sigma^2(\omega-\omega_0)^2]. \tag{3}$$

We observe that the energy of x_1 is concentrated. For instance, in the frequency interval of length $2/\sigma$

$$\omega_f = [\omega_0 - \frac{1}{\sigma}, \omega_0 + \frac{1}{\sigma}] \tag{4}$$

we have $Sx_1(\omega) > k\sigma$ with $k = 2\sqrt{\pi}/e$, and

$$\int_{\omega_f} Sx_1 = 2\pi\,\mathrm{Erf}(1) > 5\int_{\mathbb{R}\setminus\omega_f} Sx_1. \tag{5}$$

In numerical applications, time and frequency domains have a compact support for which the above computations are not exact. However, for a time-frequency domain given by $[0,T]\times[0,F]$, with ω_0 separated from zero and F, let us say $4/\sigma < \omega_0 < F - 4/\sigma$, the formulas (3)-(5) are good approximations.

In order to normalize x_1 in $L^2(0,T)$, we redefine it as $x_1 = \exp(i\omega_0 t)/\sqrt{T}$, and without loss of generality, we assume as well $\|x_2\|_{L^2} = 1$. Then, the spectrograms defined in $[0,T]\times[0,F]$ satisfy $\|Sx_i\|_{L^2} \approx 2\pi$. In addition, we observe that since x_2 is a white Gaussian noise, its power spectrum is constant and therefore Sx_2 is a uniformly distributed random variable. The spectrogram of x is

$$Sx = \lambda_1^2 Sx_1 + \lambda_2^2 Sx_2 + 2\lambda_1\lambda_2 \Re(\mathcal{G}x_1 \mathcal{G}x_2), \tag{6}$$

where $\Re(z)$ denotes the real part of z. Since Sx_1 is concentrated and large around ω_0, Sx_2 is uniformly distributed, and both have the same L^2 norm, we may expect (if $\lambda_1 \sim \lambda_2$) that setting Sx to zero below certain level curve $Sx = c$ will result in an attenuation of the noise. But not having local information on Sx_2, it is not clear how to fix c. However, we do have global information. Set $Q_f = (0,T)\times\omega_f$ and $Q_F = (0,T)\times((0,F)\backslash\omega_f)$, with ω_f given by (4). Due to the uniform distributed character of Sx_2,

$$\int_{Q_f} Sx_2 \approx 2\pi\frac{|\omega_f|}{F} = \frac{4\pi}{\sigma F} \quad \text{and} \quad \int_{Q_F} Sx_2 \approx \frac{2\pi(\sigma F - 2)}{\sigma F}. \tag{7}$$

On these sets, we also have the information on Sx_1 provided by (5). Let us consider the averages of Sx/π on the sets Q_f and Q_F given by

$$A_f = \frac{\sigma}{2\pi T}\int_{Q_f} Sx, \qquad A_F = \frac{\sigma}{\pi(\sigma F - 2)T}\int_{Q_F} Sx. \tag{8}$$

If $A_f \gg A_F$ then we may expect that the regions Q_f and Q_F will be clearly separated by some small range of level curves of S, and then it will be possible to select a cutting value, c, related to these averages. Using (5)-(7) in (8) and the inequality $|\Re(\mathcal{G}x_1\mathcal{G}x_2)| \le (Sx_1 Sx_2)^{1/2}$, we obtain

$$A_f \succeq \frac{\mathrm{Erf}(1)\sigma}{T}\lambda_1^2 + \frac{2}{TF}\lambda_2^2 - \frac{\sigma}{\pi T}\lambda_1\lambda_2\int_{Q_f}(Sx_1 Sx_2)^{1/2}, \tag{9}$$

$$A_F \preceq \frac{2(1-\mathrm{Erf}(1))\sigma}{T(\sigma F - 2)}\lambda_1^2 + \frac{2}{TF}\lambda_2^2 + \frac{2\sigma\lambda_1\lambda_2}{\pi(\sigma F-2)T}\int_{Q_F}(Sx_1 Sx_2)^{1/2}, \tag{10}$$

where the symbol $\succeq$ expresses that some terms of the left hand side are approximately equal than the corresponding of the right hand side and that the others are bigger. The cross terms in (9) and (10) may be estimated using the Hölder's inequality and our knowledge about the explicit form of Sx_1, see (3):

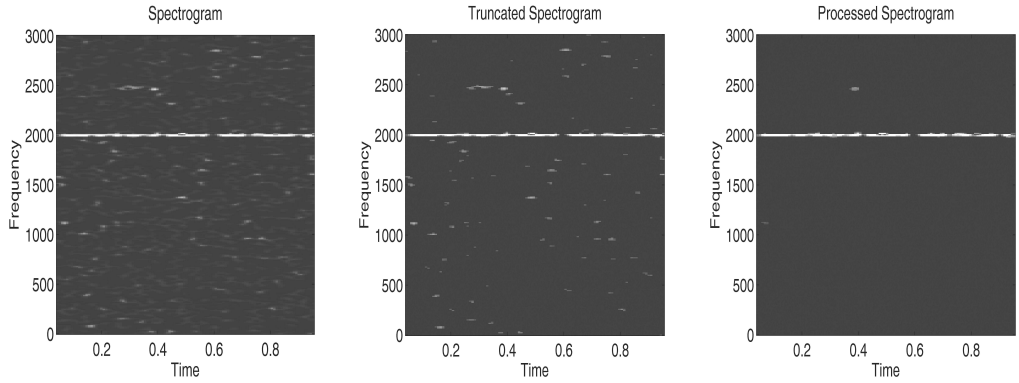

Fig. 1. Spectrogram of x, truncated spectrogram ($c = 2A_F$), and processed spectrogram. Parameter values are $\lambda_1 = 1$, $\lambda_2 = 4$, $\sigma = 1/6$, $T = 1$ and $F = 3000$.

$$\int_{Q_f} (Sx_1 Sx_2)^{1/2} \leq \Big(\int_{Q_f} Sx_1 \Big)^{1/2} \Big(\int_{Q_f} Sx_2 \Big)^{1/2} = (8\pi^2 \,\mathrm{Erf}(1)/\sigma F)^{1/2},$$
$$\int_{Q_F} (Sx_1 Sx_2)^{1/2} \leq \max_{Q_F} \sqrt{Sx_1} |Q_F|^{1/2} \Big(\int_{Q_F} Sx_2 \Big)^{1/2} = (\sigma F - 2)(2\pi c/\sigma F)^{1/2}.$$

Introducing these expressions in (9)-(10) and taking typical values of parameters such as $\sigma = 1/6$, $T = 1$ and $F = 10000$ we get

$$A_f \succeq o(10^{-1})\lambda_1^2 + o(10^{-4})\lambda_2^2 - o(10^{-3})\lambda_1\lambda_2,$$
$$A_F \preceq o(10^{-5})\lambda_1^2 + o(10^{-4})\lambda_2^2 + o(10^{-3})\lambda_1\lambda_2.$$

Therefore, if the orders of magnitude of λ_1 and λ_2 are similar and we select the cutting value on, approximately, the average value of Sx_1 in Q_f then most of the contribution of the signal x_1 to the spectrogram Sx will be conserved, being that of x_2 drastically attenuated, see Fig 1.

However, it is clear that even with a very sharp and even lucky election of the cutting value c for S and due to the random character of x_2 there will remain isolated points of magnitude bigger than c in the region $(0, T) \times ((0, F) \backslash \omega_f)$ which will perturb the modified spectrogram and the eventual procedure for instantaneous frequency estimation. In fact, this perturbation may be large if the signal to noise ratio λ_1/λ_2 is small and/or if the signal x_1 is a multi-component time varying signal, which is the case of interest. To try to smoothen this effect, our idea is to introduce an image processing based algorithm such that preserves the edges of an approximation to $(0, T) \times \omega_f$, i.e., the level set $Sx = c$, averaging the values of S in the exterior and interior of this level set. The ideal result of the application of this algorithm to the previous example is a function $\tilde{S}$ with constant value in $(0, T) \times \omega_f$ given, approximately, by (9) and a constant value in $(0, T) \times ([0, F] \backslash \omega_f)$ given, approximately, by (10). The final approximation to the spectrogram of x_1 would be given by truncating function $\tilde{S}$, see Fig 1.

3 The mathematical model

Let $x \in L^2(\mathbb{R})$ and denote u_0 to the spectrogram defined by (2) corresponding to the real, symmetric and normalized window φ. The regularity of u_0 is that inherited from the window φ, which we assume to be Lipschitz continuous. In particular, u_0 is a bounded function and then we may think of it as an *image* and consider its transformation given as the solution $u(\tau, t, \omega)$ of the following problem introduced in [9] as an edge-detection image-smoothing algorithm:

$$\frac{\partial u}{\partial \tau} - g(|G_s * \nabla u|)A(u) = \xi(u) \quad \text{in } \mathbb{R}_+ \times \Omega, \tag{11}$$

$$\nabla u \cdot \mathbf{n} = 0 \quad \text{on } \mathbb{R}_+ \times \partial\Omega, \tag{12}$$

$$u(0, \cdot, \cdot) = u_0 \quad \text{in } \Omega, \tag{13}$$

where

$$A(u) = (1 - h(|\nabla u|))\Delta u + h(|\nabla u|) \sum_{j=1,\dots,n} f_j(\frac{\nabla u}{|\nabla u|})\frac{\partial^2 u}{\partial x_j^2},$$

and the time-frequency domain $\Omega \subset \mathbb{R}^2$ is an open set that we shall consider bounded in applications: bounded in time, since the signal is of finite length, and bounded in frequency, by physical restrictions. Let us remind the properties and meaning of the terms in equation (11):

- Function G_s is a two-dimensional smoothing kernel, typically a Gaussian of variance s. The variance is a *scale parameter* which fixes the minimal size of the details to be kept in the processed image.
- Function g is non-increasing with $g(0) = 1$ and $g(\infty) = 0$. It is a *contrast* function, which allows to decide whether a detail is sharp enough to be kept.
- The combination of actions of G_s and g on ∇u rules the speed of diffusion in the evolution of the image, controlling the *enhancement* of the edges and the smoothing of noise. For points (t, ω) for which $|\nabla u|$ is large in a neighborhood of (t, ω) (being the size of the neighborhood determined by s), the value of $|G_s * \nabla u|$ is large and therefore, the diffusion coefficient $g(|G_s * \nabla u|)$ is small. Then, (for $\xi = 0$), the evolved image keeps very close to the original image at (t, ω), i.e., (t, ω) is an edge point. On the other hand, if ∇u is not large locally, then g is close to one and diffusion takes place according to the operator A.
- The diffusion operator A combines isotropic and anisotropic diffusion. The first smoothes the image by local averaging while the second enforces the diffusion only on the orthogonal direction to ∇u (along the edges). More precisely, for $\theta_j = (j-1) * \pi/n$, $j = 1, \dots, n$ we define x_j as the orthogonal to the direction θ_j, i.e., $x_j = -t \sin\theta_j + \omega \cos\theta_j$. Then, smooth non-negative functions $f_j(\cos\theta, \sin\theta)$ are designed to be *active* only when θ is close to θ_j. Therefore, the anisotropic diffusion is taken in an approximated direction

to the orthogonal of ∇u. The combination of isotropic and anisotropic diffusions is controlled by function $h(s)$, which is nondecreasing with $h(s) = 0$ if $s \leq \epsilon$, $h(s) = 1$ if $s \geq 2\epsilon$, being ϵ the *enhancement* parameter.

- Function ξ is a force term designed to accelerate the suppression of noise. It takes non-positive values for $u \leq u_* - \varepsilon$ and non-negative values for $u > u_* + \varepsilon$, for some *cutting parameter* $u_* > 0$ and a small $\varepsilon > 0$.

3.1 Mathematical properties

The following theorem is proven in [9].

Theorem 1 *Let $u_0 \in W^{1,\infty}(\Omega)$. (i) Then, for any $T > 0$, there exists a unique solution, $u \in C([0,\infty) \times \Omega) \cap L^\infty(0,T;W^{1,\infty}(\Omega))$, of problem (11)-(13). Moreover,*

$$\inf_\Omega u_0 \leq u \leq \sup_\Omega u_0 \quad \text{in } \mathbb{R}_+ \times \Omega. \tag{14}$$

(ii) Let v be a solution of problem (11)-(13) corresponding to the initial data $v_0 \in L^\infty(\Omega)$. Then, for all $T \geq 0$, there exists a constant K which depends only on $\|u_0\|_{W^{1,\infty}}$ and $\|v_0\|_{L^\infty}$ such that

$$\sup_{0\leq\tau\leq T} \|u(\tau,\cdot,\cdot) - v(\tau,\cdot,\cdot)\|_{L^\infty(\Omega)} \leq K\|u_0 - v_0\|_{L^\infty(\Omega)}. \tag{15}$$

Remark 1 *The solution ensured by this theorem is not, in general, a classical solution. The notion of solution employed in [9] is that of* viscosity solution*, which coincides with the classical solution if it is regular enough. Since we will not enter in further discussions about regularity, we refer the reader to [9,16] for technical details about this notion of solution.*

Part (*ii*) of Theorem 1 is specially useful to us for the following reason. Spectrograms of a signal are computed relative to windows, i.e, for each window a different spectrogram (image) is got. Then, the time-frequency characteristics of the signal, like instantaneous frequency, look in a slight different way if two different windows are employed. It, therefore, arises the question of stability of the final images with respect to the windows, i.e., is it possible that starting from two spectrograms of the same signal for different windows the corresponding final images are very different from each other?

Corollary 1 *Let $\varphi, \psi \in W^{1,\infty}(\mathbb{R})$ be real, symmetric and normalized windows and denote by u_0 and v_0, respectively, the corresponding spectrograms of a given signal $x \in L^2(\mathbb{R})$. Let u and v be the solutions of problem (11)-(13) corresponding to the initial data u_0 and v_0, respectively. Then*

$$\sup_{0\leq\tau\leq T} \|u(\tau,\cdot,\cdot) - v(\tau,\cdot,\cdot)\|_{L^\infty(\Omega)} \leq c\|\varphi - \psi\|_{L^2(\mathbb{R})},$$

where the constant c depends only on $\|x\|_{L^\infty}$, $\|u_0\|_{W^{1,\infty}}$, $\|u_0\|_{L^\infty}$ and $\|v_0\|_{L^\infty}$.

Proof. Let $\mathcal{G}_\eta x$ denote the Gabor's transform of x relative to the window η. The standard inequality $\Big||a|-|b|\Big| \leq |a-b|$ implies

$$\begin{aligned}\Big||\mathcal{G}_\varphi x(\omega,t)|^2 - |\mathcal{G}_\psi x(\omega,t)|^2\Big| &\leq c_1\Big||\mathcal{G}_\varphi x(\omega,t)| - |\mathcal{G}_\psi x(\omega,t)|\Big| \\ &\leq c_1|\mathcal{G}_\varphi x(\omega,t) - \mathcal{G}_\psi x(\omega,t)|,\end{aligned} \tag{16}$$

with $c_1 = ||\mathcal{G}_\varphi x(\omega,t)| + |\mathcal{G}_\psi x(\omega,t)||$. We have

$$\begin{aligned}|\mathcal{G}_\varphi x(\omega,t) - \mathcal{G}_\psi x(\omega,t)| &\leq \int_{\mathbb{R}} |x(s)(\varphi(s-t) - \psi(s-t))e^{-i\omega s}|ds \\ &\leq \|x\|_{L^2}\|\varphi - \psi\|_{L^2}.\end{aligned} \tag{17}$$

Taking the supremo in the left hand side of (16) and using (17) we obtain

$$\|u_0 - v_0\|_{L^\infty(\Omega)} \leq (\|u_0\|^{1/2}_{L^\infty(\Omega)} + \|v_0\|^{1/2}_{L^\infty(\Omega)})\|x\|_{L^2}\|\varphi - \psi\|_{L^2}. \tag{18}$$

Finally, property (15) implies the result. □

Another stability question solved with the help of Theorem 1 is whether the transformed spectrograms of two close signals, for instance of a signal x and a signal $x + \varepsilon n$, where n denotes a noise and $\varepsilon > 0$ is small, are close or not. We have the following result. Since the proof is a trivial modification of the proof of Corollary 1, we omit it.

Corollary 2 *Let $x, y \in L^2(\mathbb{R})$ be two signals and $\varphi \in W^{1,\infty}(\mathbb{R})$ be a real, symmetric and normalized window. Let u_0 and v_0 be, respectively, their spectrograms. Finally, let u and v be the solutions of problem (11)-(13) corresponding to the initial data u_0 and v_0, respectively. Then*

$$\sup_{0\leq\tau\leq T} \|u(\tau,\cdot,\cdot) - v(\tau,\cdot,\cdot)\|_{L^\infty(\Omega)} \leq c\|x-y\|_{L^2(\mathbb{R})},$$

where c depends only on $\|u_0\|_{W^{1,\infty}}$, $\|u_0\|_{L^\infty}$ and $\|v_0\|_{L^\infty}$.

4 Numerical Experiments

In this section we present numerical demonstrations of the selective smoothing edge enhancement algorithm of [9] applied to the spectrograms of synthetic and field signals. The numerical implementation of the algorithm starts with a pre-processing of the signal. Field data recorders are set to 44.1 KHz meanwhile wolves signals are rarely out of the range $200 - 3000$ Hz, so we start by

filtering and downsampling the signal to speed up computations. We then pass to the computation of the spectrogram by applying the discrete fast fourier transform (dfft) to the convolution of the signal with the window. The dfft is evaluated in time intervals of size 2^w, with the *width*,w, usually in the range $8 - 12$. To obtain an image as *continuous* as possible, the time intervals are overlapped according to the value of $p \in (0, 1)$. In each of these intervals, we perform the convolution of the signal with a normalized discrete gaussian window with support on $(-t_0/2, t_0/2)$ and variance $\sigma = t_0/6$, where t_0 is the size of the time intervals. The size, t_0, is an increasing function of w, implying that for larger values of w we get a better frequency resolution in the spectrogram, and therefore, a poorer time resolution, as a consequence of the Heinsenberg's Principle.

Once the spectrogram is produced, it is normalized in the usual digital image range $[0, 255]$, obtaining in this way the initial datum for problem (11)-(13). We use a time explicit Euler scheme with finite differences in space to find the numerical approximation of the solution, u. We follow the discretization indicated in [9], considering four directions based in a nine nodes local grid to implement the anisotropic diffusion operator, and use these nodes to compute also the discretized laplacian and gradient, according to the formulas given in [9]. The convolution $G_s * \nabla u$ is done in a nine nodes local grid, with variance $s = 1$, and normalized to get a partition of the unity in the discrete grid. Since the distance between time-frequency nodes is set equal to one, we choose constant evolution steps $d\tau < 0.5$, for stability issues. The force term function, ξ, is defined to enforce either exponential increase or exponential decrease of u depending on its relative value with respect to some positive constant, μ, which in the experiments we fix as $\mu = \int_\Omega u$. This value is motivated by the discussion in Section 2. More precisely, we use the function

$$\xi(s) = \begin{cases} -s/\mu & \text{if } s \leq \mu - 1/\mu, \\ s(s-\mu) & \text{if } |s - \mu| < 1/\mu, \\ s/\mu & \text{if } s \geq \mu + 1/\mu. \end{cases}$$

Finally, since the scheme is explicit in time, it can not be expected the discrete comparison principle to hold, see (14). We therefore re-normalize the solution in each time iteration to the range $[0, 255]$.

Summarizing, the parameters in the model come from three sources: the spectrogram definition, the image processing partial differential equation, and its numerical implementation. From the spectrogram we get the variance of the gaussian window, σ, which is determined by the width, w, and the overlapping, p. From the PDE we have the enhancement parameter, ϵ, the scale parameter (variance of the smoothing kernel), s, and the normalizing parameter, μ. Finally, from the discretization we have the evolution step, $d\tau$, and the

number of advances or iterations, k. In the experiments, we keep fixed those parameters which seems to be less sensible. More precisely, we always take

$$p = 0.99, \quad \epsilon = \frac{1}{2}\max|\nabla u_0|, \quad s = 1, \quad \mu = \int_\Omega u, \quad d\tau = 0.1.$$

Hence, the only parameters we play with in the experiments are the width and the number of iterations. Both of them are very related with the computer execution time since the width determines the time-frequency grid size. It is not clear how to fix them a priori. On one hand, the width is related to the smoothness of the discrete spectrogram and variations of this parameter may induce breaks in the lines of instantaneous frequencies, among other effects. Similarly, when the number of iterations increases the image gets more and more diffused making possible that some not very neatly defined edges may disappear.

Finally, to show more clearly the advantages of our technique, in the subsequent plots we used a simple algorithm to produce candidates to instantaneous frequency lines of the corresponding spectrograms. Let $\Omega = [0,T] \times [0,F]$ be the time-frequency domain of the image and $u : \Omega \to [0,255]$ be the starting image. We consider its truncation

$$v(t,\omega) = \begin{cases} u(t,\omega) & \text{if } u(t,\omega) \geq \beta \\ 0 & \text{elsewhere,} \end{cases}$$

with $\beta = \text{Mean}_\Omega(u)$ in the experiments. For each $t \in [0,T]$ we consider the N connected components of the set $\{\omega \in (0,F) : v(t,\omega) > 0\}$, say $C_n(t)$, for $n = 1,\ldots,N(t)$, and define the function

$$\text{IF}(t,n) = \frac{\int_{C_n(t)} \omega v(t,\omega) d\omega}{\int_{C_n(t)} v(t,\omega) d\omega},$$

which is the frequency gravity center of the component $C_n(t)$. In this way, we shrink each connected component to one point to which we assign the average image intensity through the function $\text{INT}(t,n) = \text{Mean}_{C_n(t)}(v(t,\cdot))$. Finally, we plot function IF only for components with averaged intensity, INT, greater than a certain threshold, $i \in [0,255]$, which we take in the experiments as low as $i = 3$. This final image does not seem to be very sensible under small perturbations of the parameters β and i.

Experiment 1. We used a recording done in captivity, see [17], from where we extracted a very clean signal of app. 0.7 seconds containing the howls of two wolves. We filtered the signal to the range $200 - 1200$ Hz and added a synthetic white uniform noise with a signal to noise ratio (SNR) equal to one. The width was set to 9 and we run 30 iterations of the algorithm. In the first row of Fig. 2 we plot, respectively, the spectrogram of the signal with

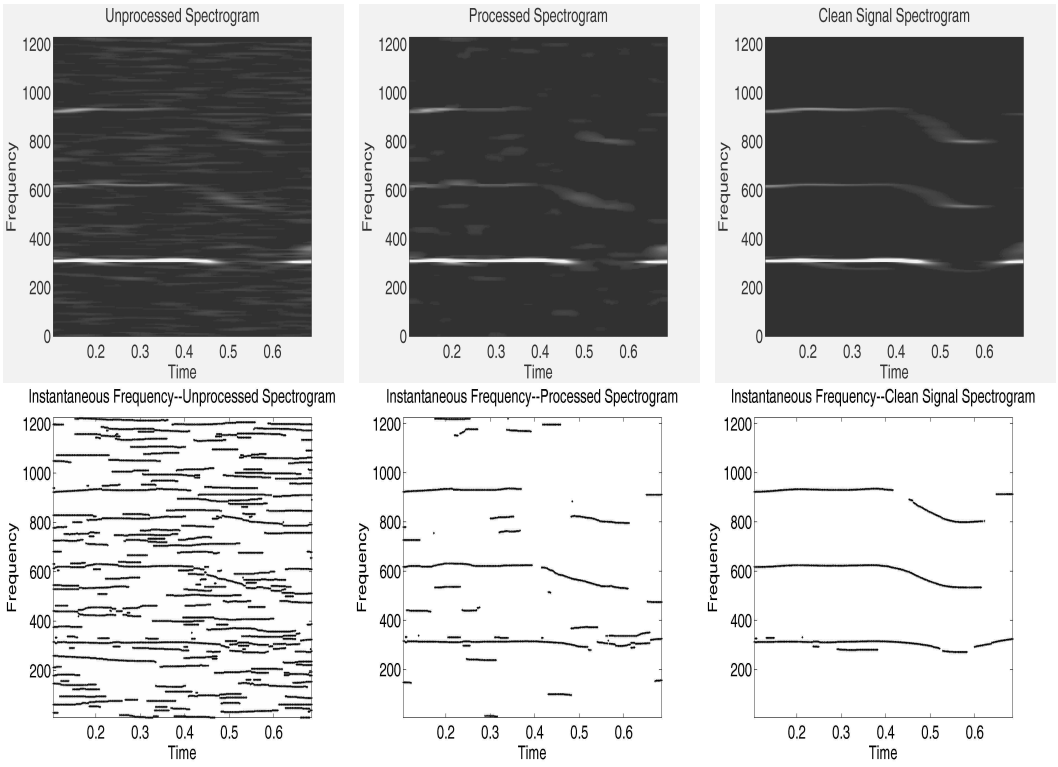

Fig. 2. Spectrograms and IF lines corresponding to Experiment 1, of a clean signal containing two howls (one with two harmonics), corrupted with an artificial white noise.

the added noise (initial datum), the processed spectrogram and the spectrogram of the original clean signal. In the second row we plot the corresponding instantaneous frequency lines computed with a simple maxima localization algorithm. Although the second harmonic of the less intense howl is broken, we see that the instantaneous frequency lines are much better identified from the processed spectrogram than from the original noisy spectrogram.

Experiment 2. We used a recording done in wilderness, see [18], from where we extracted a signal of app. 0.55 seconds which is affected by a strong background noise. The width was set to 10 and we run 200 iterations of the algorithm. Since the width is larger than in the previous experiment, the spectrogram is stretched in the frequency direction, producing a more diffused image. In the first row of Fig. 3 we plot the spectrogram of the original signal (initial datum) and the processed spectrogram, and in the second row, the corresponding instantaneous frequency lines are plotted. We identify three possible howls, one with two harmonics in the approximated steady frequencies 400 and 800 Hz, another in about 600 Hz (decreasing in time), and finally, another with two harmonics starting at 1000 (decreasing) and 700 Hz, respectively, although the latter becomes too weak to be detected, after a while.

Experiment 3. In this experiment we used a one second 6KHz synthetic

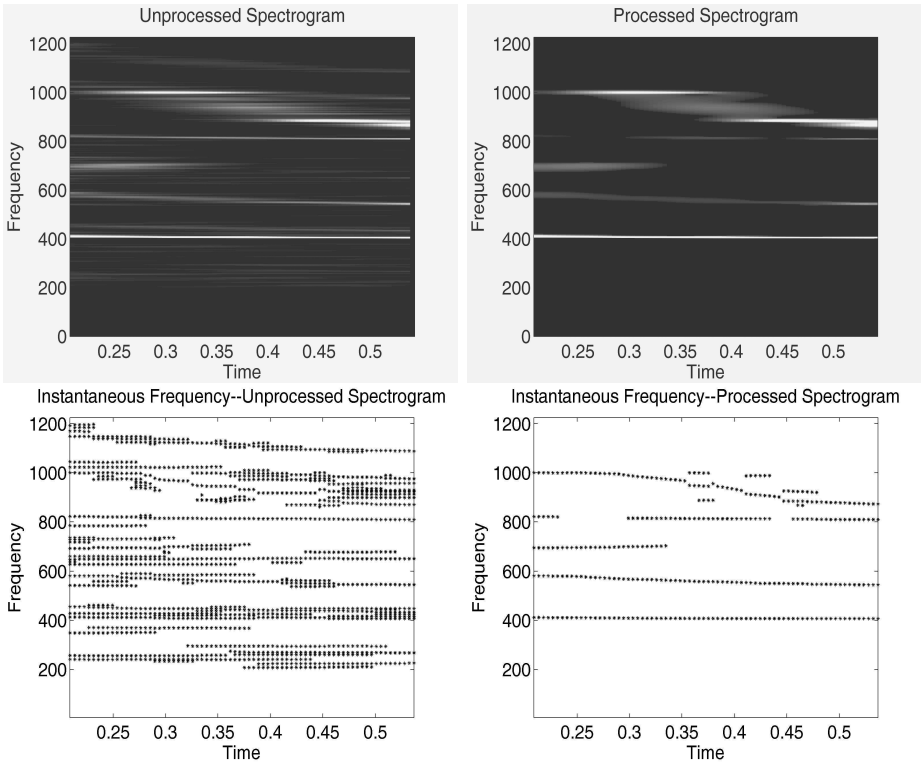

Fig. 3. Spectrograms and IF lines corresponding to Experiment 2, of a very noisy field recorded signal containing, apparently, three howls (two with two harmonics).

signal composed by the addition of two signals. The first is the addition of pure tones and chirps:

$$x_1(t) = c_1(\sin 2\pi 1000t + \sin 2\pi 1100t + \sin 2\pi 1300t^2 + \sin 2\pi 800t^3),$$

while the second, x_2, is a uniformly distributed real random variable. We normalize them to have $\|x_i\|_{L^2} = 1$ (so the constant c_1) and define the test signal as $x = x_1 + 4x_2$, i.e., with SNR $= 4$. The width is set to 10 and the number of iterations to 50. As in previous figures, in the first row of Fig. 4 we plot the spectrogram of the signal with the added noise, the processed spectrogram and the spectrogram of the original clean signal, and in the second row, the corresponding instantaneous frequency lines. It is interesting to observe that even for very close instantaneous frequency lines, the processed spectrogram keeps the separation, despite being produced by a diffusive transformation of the noisy signal. We again notice the large qualitative difference between the instantaneous frequency lines detection of the noisy and the processed image, plotted in the second row.

Acknowledgements. This work was motivated by the enthusiastic interest of the project biologists Luis LLaneza and Vicente Palacios, to whom we thank their efforts to provide us with an increasingly improved sequence of wolves chorus recordings, obtained always in uncomfortable conditions and often in risky situations.

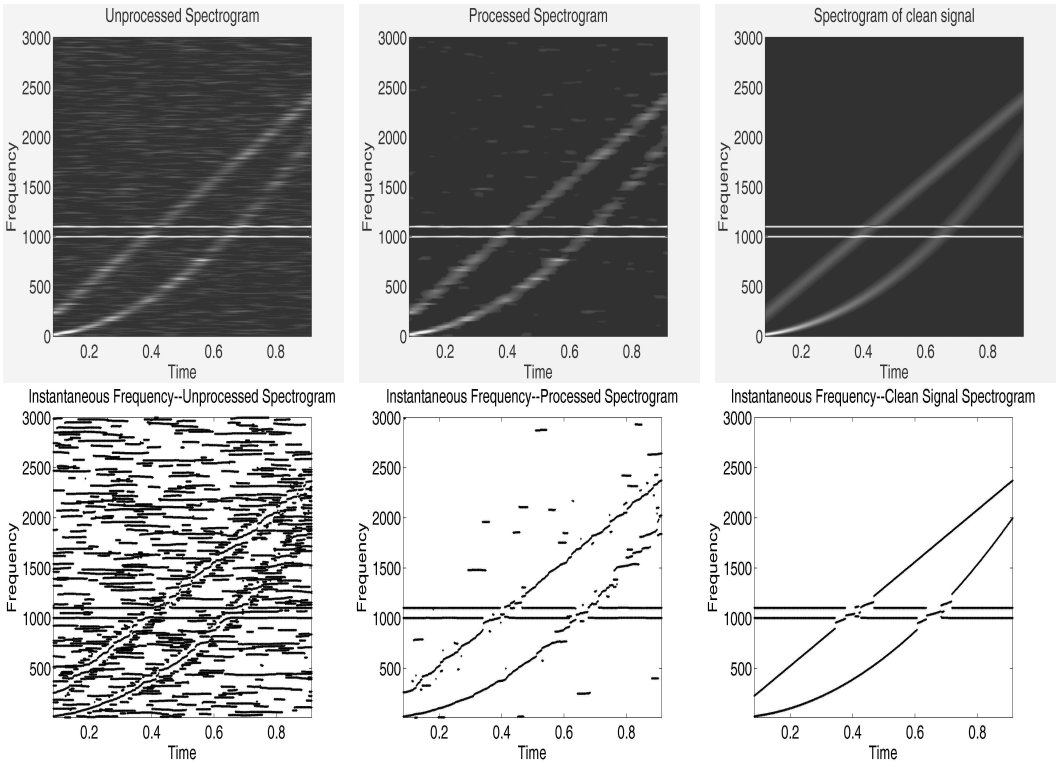

Fig. 4. Spectrograms and IF lines corresponding to Experiment 3, of a synthetic signal with $SNR = 4$.

References

[1] A. Skonhoft, The costs and benefits of animal predation: An analysis of Scandinavian wolf re-colonization, Ecological Economics 58 4 (2006) 830–841.

[2] L. R. Rabiner, R. W. Schafer, Digital processing of speech signals, Prentice-Hall, New Jersey, 1978.

[3] S. Mallat, A wavelet tour of signal processing, Academic Press, London, 1998.

[4] S. Chandra Sekhar, T. V. Sreenivas, Adaptive spectrogram vs. adaptive pseudo-WignerVille distribution for instantaneous frequency estimation, Signal Processing, 83 7 (2003) 1529–1543.

[5] I. Djurović, L. Stanković, An algorithm for the Wigner distribution based instantaneous frequency estimation in a high noise environment, Signal Processing, 84 3 (2004) 631–643.

[6] G. Viswanath and T. V. Sreenivas, A IF estimation using higher order TFRs, Signal Processing, 82 2 (2002) 127–132.

[7] A.G. Poulimenos, S.D. Fassois, Parametric time-domain methods for non-stationary random vibration modelling and analysis A critical survey and comparison, Mechanical Systems and Signal Processing, 20 4 (2006) 763–816.

[8] H. Zou, D. Wang, X. Zhang, Y. Li, Nonnegative time-frequency distributions for parametric time-frequency representations using semi-affine transformation group, Signal Processing, 85 9 (2005) 1813–1826.

[9] L. Álvarez, P. L. Lions, J. M. Morel, Image selective smoothing and edge detection by nonlinear diffusion. II, SIAM J. Numer. Anal, 29 3 (1992), 845-866.

[10] G. Aubert, P. Kornprobst, Mathematical problems in image processing, Applied Mathematical Sciences, Springer, New York, 2002.

[11] G. Sapiro, Geometric partial differential equations and image analysis, Cambridge University Press, Cambridge, 2001.

[12] H. Yurk, L. Barret-Lennard, J.K.B. Forf, C.O. Matkins, Cultural transmission within maternal lineages:vocal clans in resident killer whales in southern Alaska, Animal Behavior, 63 (2002) 1103–1119.

[13] S. Datta, C. Sturtivant, Dolphin whistle classification for determining group identities, Signal Processing, 82 2 (2002) 15–22.

[14] A. Searby , P. Jouventin, T. Aubin, Acoustic recognition in macaroni penguins: an original signature system, Animal Behaviour, 67 4 (2004) 615–625.

[15] G. Jones, E. C. Teeling, The evolution of echolocation in bats, Trends in Ecology & Evolution, 21 3 (2006) 149–156.

[16] M. G. Crandall, H. Ishii, and P. L. Lions, User's guide to viscosity solutions of second order partial differential equations, Bull. Amer. Math. Soc. (N.S.), 27 1 (1992) 1–67.

[17] L. LLaneza, V. Palacios, Asesores en Recursos Naturales, S.L., Field recordings obtained in wilderness in Asturias (Spain) in the 2003 campaign.

[18] L. LLaneza, V. Palacios, Asesores en Recursos Naturales, S.L., Field recordings obtained in captivity in Spain and Portugal in the 2005 campaign.