

Disentangling Modes and Interference in the Spectrogram of Multicomponent Signals

Kévin Polisano*, Sylvain Meignen*, Nils Laurent†, Hubert Leterme‡

* Univ. Grenoble Alpes, CNRS, F-38000 Grenoble, France

† Ens de Lyon, CNRS, Laboratoire de Physique, F-69342 Lyon, France

‡ Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR 6072, F-14000 Caen, France

Abstract—In this paper, we investigate how the spectrogram of multicomponent signals can be decomposed into a mode part and an interference part. We explore two approaches: (i) a variational method inspired by texture-geometry decomposition in image processing, and (ii) a supervised learning approach using a U-Net architecture, trained on a dataset encompassing diverse interference patterns and noise conditions. Once the interference component is identified, we explain how it enables us to define a criterion to locally adapt the window length used in the definition of the spectrogram, for the sake of improving ridge detection in the presence of close modes. Numerical experiments illustrate the advantages and limitations of both approaches for spectrogram decomposition, highlighting their potential for enhancing time-frequency analysis in the presence of strong interference.

Index Terms—Time-frequency, multicomponent signal, interference, image decomposition, variational method, deep learning, signal reconstruction, ridge detection, frequency estimation.

I. INTRODUCTION

NON-stationary signals such as audio (music, speech, bird songs) [1], electrocardiogram [2] or thoracic and abdominal movement signals [3] can be modelled as a superimposition of amplitude and frequency-modulated (AM/FM) modes, referred to as *multicomponent signal* (MCS), and defined as

$$f(t) = \sum_{p=1}^P f_p(t), \text{ with } f_p(t) = A_p(t)e^{2i\pi\phi_p(t)}, \quad (1)$$

where the *instantaneous amplitudes* (IAs) $A_p(t)$ and the *instantaneous frequencies* (IFs) $\phi_p'(t)$ are set to be positive. This representation is associated with the ideal *time-frequency* (TF) signature defined as:

$$IT_f(t, \eta) = \sum_{p=1}^P A_p(t)\delta(\eta - \phi_p'(t)), \quad (2)$$

where δ is the Dirac distribution. To estimate IT_f , the *short-time Fourier transform* (STFT) is commonly used

$$V_f^h(t, \eta) = \int_{\mathbb{R}} f(x)h(x-t)e^{-i2\pi\eta(x-t)}dx, \quad (3)$$

where h is a real-valued window. The squared modulus of the STFT, known as the *spectrogram*, is given by $S_f^h := |V_f^h|^2$.

The ideal TF signature associated with the spectrogram is the same as (2) replacing the amplitudes with their squared values.

Traditionally, IFs are estimated on the spectrogram by considering *ridge detection* (RD), which involves identifying and connecting local maxima along the frequency axis over time [4]. In connection with IF estimation, significant efforts have been made to enhance the readability of TF representations, including reassignment methods [5] and the *Fourier-based synchrosqueezing transform* (FSST) [6]. However, all these techniques become irrelevant in the presence of strong interference in the TF plane. Indeed, when two or more modes are close or intersect in the TF plane, reassignment operators become unreliable, while a mode may no longer be associated with a single spectrogram ridge. The resulting phase mixing leads to constructive or destructive interference, creating complex patterns that are difficult to analyze [7]. More precisely, the spectrogram of a MCS f , as defined in (1), is the sum of the spectrograms of each mode plus undesirable extra cross-terms responsible of the interference:

$$|V_f^h|^2 = \sum_{p=1}^P |V_{f_p}^h|^2 + 2 \sum_{p=1}^P \sum_{\substack{q=1 \\ q \neq p}}^P \text{Re}(V_{f_p}^h V_{f_q}^{h*}). \quad (4)$$

Thus, to have an insight into where the interference term can be neglected is essential for IF estimation for instance. In this regard, several recent studies have aimed at localizing interference in the TF plane using techniques such as hyperbolic spline fitting [8] or neural network [9]. The latter work primarily focused on segmenting interference regions using a convolutional neural network (CNN)-based classification approach. However, the dataset is limited to linear chirps, does not account for complex interference patterns that may arise, and relies on a single window choice. Moreover, this study neither addresses the separation of interference from the rest of the signal, nor proposes strategies for IF estimation in regions affected by interference. We shall mention that as the spectrogram is constrained by the choice of a window, the Wigner-Ville distribution (WVD) offers an alternative for IF estimation. For that purpose, one needs to separate the interference part from the rest of the signal as proposed in [10]. However, in the spectrogram representation interference are mitigated by windowing, but, in the WVD representation of a MCS, interference between any two modes remain, leading

to highly complex patterns when more than two modes are present (see *e.g.* [10, Fig. 12]).

Our goal in the present paper is to stick to the spectrogram representation and investigate how it can be decomposed into a mode part and an interference part. To this end, we propose to consider two different paths. The first one consists in revisiting variational approaches involving proximal techniques that were initially proposed to separate texture from geometry in images [11]. In our context, we interpret the “geometrical” part of the spectrogram, viewed as an image, as the so-called “mode part”, while the “texture” part corresponds to interference in the spectrogram. The second approach we consider relies on supervised learning, employing a U-Net architecture [12] trained on a dataset designed to capture diverse interference scenarios. Our dataset also incorporates noise to enhance robustness. Once the interference component has been extracted from the spectrogram, we leverage this information to refine RD using an adaptive optimal window selection for improved spectrogram-based IF estimation.

The paper is organized as follows. In Sections II and III, we detail the two different techniques we propose to separate the mode parts from the interference part in the spectrogram. Section IV presents numerical experiments illustrating the advantages and limitations of the proposed methods, along with insights into how interference extraction can benefit ridge estimation.

II. FIRST MODEL BASED ON A VARIATIONAL APPROACH

A. Oscillating Interference Patterns on Toy Models

In the case of two parallel harmonics, a simple analytical expression for the spectrogram can be derived. Considering the sum of two harmonics, $f(t) = A_1 e^{i2\pi f_1 t} + A_2 e^{i2\pi f_2 t}$ and computing its STFT with a Gaussian window $h_\sigma(x) = \frac{1}{\sigma} e^{-\frac{x^2}{\sigma^2}}$, we obtain

$$|V_f^{h_\sigma}(t, \eta)|^2 = \underbrace{A_1^2 e^{-2\pi\sigma^2(\eta-f_1)^2} + A_2^2 e^{-2\pi\sigma^2(\eta-f_2)^2}}_{\text{Mode part}} + \underbrace{2A_1 A_2 e^{-\pi\sigma^2((\eta-f_1)^2 + (\eta-f_2)^2)} \cos(2\pi(f_2 - f_1)t)}_{\text{Interference part}}, \quad (5)$$

where the so-called “Mode part” (first two terms) represents the individual modes, while the “Interference part” (last term) corresponds to the interference pattern. This interference strongly depends on the window length parameter σ [7]. It can be observed that the interference term oscillates along the time axis with a frequency $\delta f := |f_2 - f_1|$ along the time axis, while its amplitude depends only on frequency.

As an example, we compute the spectrogram of two harmonics separated by $\delta f = 40\text{Hz}$. The first harmonic is set at 115Hz, while the second at 155Hz. We consider a signal of length $N = 256$ over the interval $[0, 1]$ with N frequency-bins and a window size of $\sigma = 0.01$. The resulting spectrogram, along with the decomposition in “Mode part” and “Interference part”, are shown in Fig. 1. Note that, one can obtain analytically the same kind of oscillating pattern for

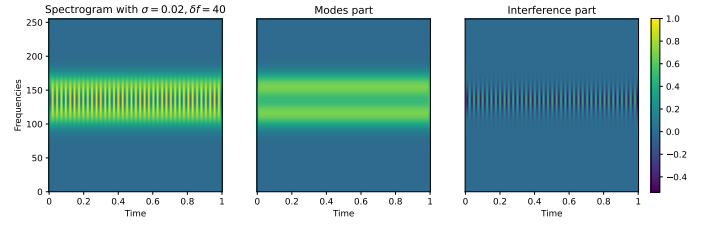


Fig. 1. Spectrogram of two harmonics (5) separated by $\delta f = 40\text{Hz}$ (left), along with its decomposition into a “Mode part” (middle) and an “Interference part” (right).

the interference term in the case of two parallel linear chirps [7].

B. TV-minimization using the G -norm

A spectrogram containing mode parts and interference part can be viewed as a geometric part (mode parts) overlaid with a texture part (interference part). Inspired by image decomposition algorithms that separate an image into texture and cartoon part [13], we propose to transpose this approach to the case of spectrograms containing interfering modes. In [13], the geometric part is modeled as a function of bounded variations, denoted by u , while the texture, denoted by v , belongs to a space of oscillatory functions. The spectrogram, denoted by s , should then be decomposed as $s \approx u + v$. For the sake of clarity, we briefly summarize this approach. Assuming that the discrete spectrogram s is represented as an $N \times N$ image, we define $\mathcal{X} = \mathbb{R}^{N \times N}$ and $\mathcal{Y} = \mathcal{X} \times \mathcal{X}$. The discrete total variation for any $\mathbf{u} \in \mathcal{X}$ is given by

$$J(\mathbf{u}) = \sum_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}} \|(\nabla \mathbf{u})_{i,j}\|_2, \quad (6)$$

where the subscript 2 denotes the Euclidian norm on \mathcal{Y} . To model the oscillations, the subspace

$$\mathcal{G} = \{\mathbf{v} \in \mathcal{X}, \exists \mathbf{g} \in \mathcal{Y} \text{ s.t. } \mathbf{v} = \text{div}(\mathbf{g})\}, \quad (7)$$

is introduced, where div denotes the discrete divergence. This space can be equipped with the norm:

$$\|\mathbf{v}\|_{\mathcal{G}} = \inf_{\mathbf{g} \in \mathcal{Y}, \mathbf{v} = \text{div}(\mathbf{g})} \left\{ \|\mathbf{g}\|_{\infty} = \max_{i,j} \|g_{i,j}\|_2 \right\}. \quad (8)$$

\mathcal{G}_μ is then defined as the set of elements in \mathcal{G} with norm smaller than μ . The decomposition of the spectrogram image s into a bounded variation part \mathbf{u} and a texture part \mathbf{v} is performed by solving the minimization problem [14]:

$$\underset{(\mathbf{u}, \mathbf{v}) \in \mathcal{X} \times \mathcal{G}_\mu}{\text{argmin}} \frac{1}{2} \|\mathbf{s} - \mathbf{u} - \mathbf{v}\|_2^2 + \lambda J(\mathbf{u}), \quad (9)$$

the subscript 2 corresponding to the Euclidian norm on \mathcal{X} .

C. Implementation with Alternate Projections

Fixing \mathbf{u} , one minimizes (9) by considering:

$$\inf_{\mathbf{v} \in \mathcal{G}_\mu} \|\mathbf{s} - \mathbf{u} - \mathbf{v}\|_2^2. \quad (10)$$

Fixing \mathbf{v} , one minimizes (9) as:

$$\operatorname{argmin}_{\mathbf{u} \in \mathcal{X}} J(\mathbf{u}) + \frac{1}{2\lambda} \|\mathbf{s} - \mathbf{u} - \mathbf{v}\|_2^2. \quad (11)$$

Eq. (10) is equivalent to projecting $\mathbf{s} - \mathbf{u} - \mathbf{v}$ onto \mathcal{G}_μ which is performed using Chambolle's algorithm [15], while the solution to (11) is $\mathbf{u} = \mathbf{s} - \mathbf{v} - \pi_{\mathcal{G}_\mu}(\mathbf{s} - \mathbf{v})$ (where π denotes the projection) [15]. Such a decomposition is summarized in Algorithm 1.

Algorithm 1: Image separation algorithm [14]

- 1: **Initialization:** $\mathbf{u}_0 = \mathbf{s}$, $\mathbf{v}_0 = \mathbf{v}_1 = \mathbf{u}_1 = 0$, ϵ
 - 2: $k = 0$
 - 3: **while** $\max(\|\mathbf{u}_{k+1} - \mathbf{u}_k\|_\infty, \|\mathbf{v}_{k+1} - \mathbf{v}_k\|_\infty) > \epsilon$ **do**
 - 4: $\mathbf{v}_{k+1} = \pi_{\mathcal{G}_\mu}(\mathbf{s} - \mathbf{u}_k)$
 - 5: $\mathbf{u}_{k+1} = \mathbf{s} - \mathbf{v}_{k+1} - \pi_{\mathcal{G}_\lambda}(\mathbf{s} - \mathbf{v}_{k+1})$
 - 6: **end while**
-

Note that with such an approach, the noise, if any, is captured in the texture part of the image. It is worth noting here that developments on textured images denoising were proposed based on soft-thresholding, assuming the added noise is Gaussian white (see [14], for instance). However, in the case of spectrograms of noisy signals, the noise is filtered and cannot be removed by soft-thresholding as proposed in [14].

III. SECOND MODEL BASED ON DEEP LEARNING

A. A Supervised Approach Based on U-Net

If the MCS consists of two modes corrupted by some noise, denoted as $\tilde{f} = f_1 + f_2 + \epsilon$, its spectrogram reads:

$$|V_{\tilde{f}}^h|^2 = |V_{f_1}^h|^2 + |V_{f_2}^h|^2 \quad (12)$$

$$+ |V_\epsilon^h|^2 \quad (13)$$

$$+ 2\operatorname{Re}(V_{f_1}^h V_{f_2}^{h*}) \quad (14)$$

$$+ 2\operatorname{Re}(V_{f_1}^h V_\epsilon^{h*}) + 2\operatorname{Re}(V_{f_2}^h V_\epsilon^{h*}), \quad (15)$$

where (12) corresponds to the ideal spectrogram of the two modes (free of noise and interference), (13) to the noise spectrogram, (14) to the interference between modes, and (15) to the interference between the noise and the modes.

The objective is to train a neural network that, given a noisy spectrogram $|V_{\tilde{f}}^h|^2$, can extract both the ideal spectrogram (12) and the interference (14), while discarding both the noise spectrogram (13) and the interference components due to noise (15). A U-Net model [12] is used as the backbone due to its ability to capture both local and global features through skip connections. An overview of the U-Net decomposition is shown in Fig. 2.

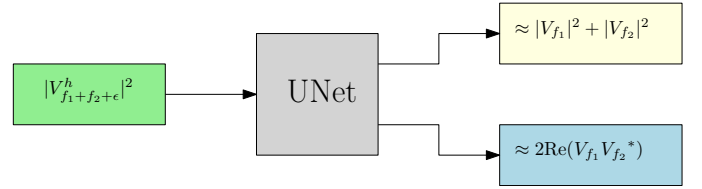


Fig. 2. Illustration of the U-Net decomposition of the spectrogram (12)-(14).

B. Synthetic Data Generation with IF and IA Splines

In our approach, we consider MCSs in which both the instantaneous frequencies $\phi'_p(t)$ and amplitudes $A_p(t)$ of the modes are generated using cubic splines. For each component, we define in the TF plane a set of control points $\{t_\ell, \eta_\ell\}_{\ell=1}^L$, uniformly distributed in time and randomly sampled in frequency within a predefined range. The instantaneous frequency $\phi'_p(t)$ is thus modeled as a cubic spline function interpolating the values $\{\eta_\ell\}_{\ell=1}^L$ at the knots $\{t_\ell\}_{\ell=1}^L$, to ensure smooth variations while preventing unrealistic abrupt changes. The phase function $\phi_p(t)$ is then obtained by integration. Similarly, the amplitude function $A_p(t)$ is generated using another spline interpolation, ensuring smooth energy variations over time. We then generate $P = 2$ modes, defined as $f_p(t) = A_p(t)e^{2i\pi\phi_p(t)}$ and construct the signal $f(t) = f_1(t) + f_2(t)$ of length $N = 256$, uniformly sampled over $[0, 1]$. Finally, we compute the STFT of f using a Hann window h of length w , yielding the corresponding spectrogram image $|V_f^h|^2$. In practice, we use $L = 5$ control points for the amplitudes and instantaneous frequencies, producing sufficient curves variability. Moreover, we ensure that the curve remains within the frequency interval $[\Delta, N - 2\Delta]$, with $\Delta = \frac{2N}{w}$, where N is the length of the generated signal and the number of frequency-bins, so as to avoid aliasing. A Gaussian white noise ϵ is added to the signal f , the input SNR being randomly chosen for each image, following a uniform distribution between 0 and 20. The Hann window size w is also randomly selected between 10 and 80, and the resulting spectrogram image is of size $N \times N$. For each noiseless spectrogram $\mathbf{s} = |V_f^h|^2$ of two interfering modes, the terms (12)-(14) are computed and denoted by \mathbf{u} and \mathbf{v} , respectively. The U-Net input is $\tilde{\mathbf{s}} = |V_{f+\epsilon}^h|^2$, representing the spectrogram of the noisy signal, while the label output (\mathbf{u}, \mathbf{v}) consists of a two-channel ground truth containing the ideal modes spectrogram (12) and the interference term (14). This procedure is repeated to generate a dataset of $M = 10^5$ triplet spectrogram images $\{\tilde{\mathbf{s}}_m, \mathbf{u}_m, \mathbf{v}_m\}_{m=1}^M$.

C. Implementation Details

These $M = 10^5$ samples are then split into training and test datasets. The number of downsampling/upsampling levels is set to 3. Mean Squared Error (MSE) is used as loss function, and the Adam optimizer is employed with a learning rate of $5 \cdot 10^{-4}$ and a weight decay of $5 \cdot 10^{-8}$. Training is performed for 40 epochs, and to enhance generalization, random cropping is applied to training images as a data augmentation technique.

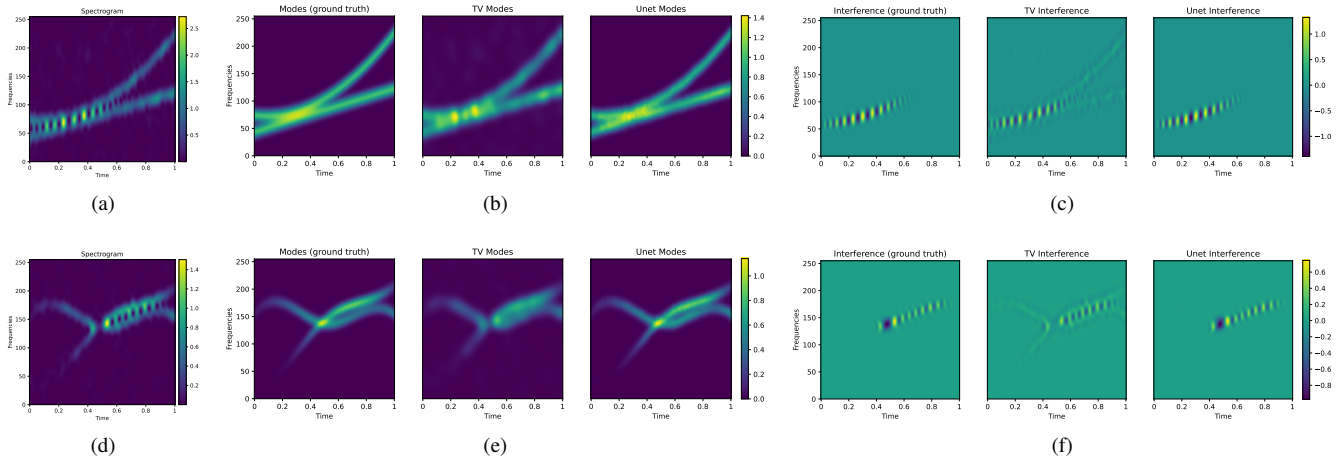


Fig. 3. Spectrograms of a noisy signal consisting of (a) the superposition of a linear and hyperbolic chirps ($w = 20$, $\text{SNR} = 10$) and (d) of two spline modes, generated by the procedure described in Section III-B ($w = 23$, $\text{SNR} = 10$). (b)-(e) Modes parts and (c)-(f) Interference parts in the spectrogram decomposition, presented in the following order: ground truth, TV-based decomposition, and U-Net predictions.

IV. RESULTS AND DISCUSSIONS

This section presents two different types of numerical experiments. We first propose a comparison between the variational and UNet approaches regarding mode and interference parts separation on spectrograms of MCSs. Secondly, we investigate how this separation can be used to estimate the IFs of the modes. All the experiments conducted in Python can be reproduced using the Jupyter notebook [16].

A. Separating Mode parts from Interference Part

The comparison between the approaches introduced in Sec. II and III is made on noisy spectrogram examples with an input SNR of 10 dB, shown in Fig. 3 (a)-(d). Fig. 3a corresponds to the spectrogram of a signal composed of the superimposition of a linear chirp and a hyperbolic chirp, whose closely spaced modes interfere (using a Hann analysis window of size $w = 20$). Fig. 3d represents the spectrogram of two interfering spline modes, generated using the procedure described in Section III-B (with $w = 23$), which was not included in the training or test dataset. By construction, we have access to *ground truth* (GT) images (\mathbf{u}, \mathbf{v}) corresponding to the spectrograms of the individual modes (12) and the interference resulting from the cross terms (14). GT images (\mathbf{u}, \mathbf{v}) are shown on the left of each sub-figure in Fig. 3 (b)-(c)-(e)-(f). The effectiveness of the two approaches in separating the mode part from the interference part is evaluated with respect to GT images.

In Algorithm 1 we set $\lambda = 1$, $\mu = 100$, $\epsilon = 10$ and the number of iterations in Chambolle's algorithm to perform the projection π is set to 100. The inputs are the spectrograms $\tilde{\mathbf{s}}$ in Fig. 3 (a)-(d), and the outputs ($\mathbf{u}_{\text{TV}}, \mathbf{v}_{\text{TV}}$) are displayed in the middle of each sub-figure of Fig. 3 (b)-(c)-(e)-(f), while the U-Net predictions ($\mathbf{u}_{\text{unet}}, \mathbf{v}_{\text{unet}}$) appear to their right. It can be observed that Algorithm 1 successfully captures the interference in the texture component (along with associated noise), but some geometric structures related to the modes

remain. Moreover, the TV minimization term produces a somehow diffuse estimation of the mode part. In contrast, the U-Net predictions are closer to GT images (\mathbf{u}, \mathbf{v}): the interference term is accurately captured and denoised, while the geometric component preserves the mode spreading, albeit with some irregularities. Details about the evolution of the loss function during the training and of the PSNR between prediction and GT images can be found in the notebook [16].

It is worth noting that the network is trained on spectrogram data generated with analysis windows of varying sizes, demonstrating its adaptability. By fixing the analysis window size, the predictions could be improved, and irregularities could be reduced. Similarly, by prioritizing training images with high-energy interference patterns, the number of required training samples could be significantly reduced, at the cost of reduced adaptability. Indeed, in many cases, the interference between the two spline modes is weak, either because they do not cross or are not sufficiently close, or because the window size is too large to induce significant interference.

B. Adapting the Window Length Based on Interference Patterns

In this section, we investigate how to select an appropriate window length based on the proposed decomposition into mode and interference components. We do not explore adaptive window selection in both time and frequency, but this could be easily achieved using the framework developed hereafter. Let $(\mathbf{u}^{(w)}, \mathbf{v}^{(w)})$ denote the decomposition in mode and interference components obtained from the trained U-Net applied to the initial noisy spectrogram image $V_{\tilde{f}}^{h(w)}$, where $h(w)$ represents the Hann window of length w . Our strategy for determining an optimal window length involves first defining a collection of candidate windows, $\mathbf{w} = \{w_q, q = 1, \dots, Q\}$, leading, for each q , to different decompositions $(\mathbf{u}^{(w_q)}, \mathbf{v}^{(w_q)})$ for each $\tilde{\mathbf{s}}^{(w_q)} = V_{\tilde{f}}^{h(w_q)}$. Then, for each time index n , we estimate locally the index

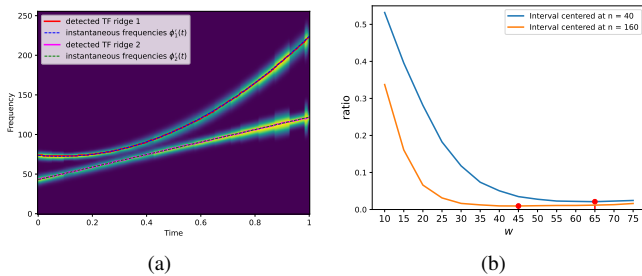


Fig. 4. (a) V_{adap} with detected TF ridges, (b) evolution of the interference-to-signal ratio as a function of window size w at time indices $n = 40$ and $n = 160$ (resp. $t_n = 0.15$ and $t_n = 0.62$ on the time axis), with the minimum ratio achieved at \hat{q}_n (16) highlighted in red.

q associated with the minimum relative contribution of the interference to the noisy spectrogram, namely:

$$\hat{q}_n = \underset{q}{\operatorname{argmin}} \frac{\sum_{m \in \mathcal{V}_n, k} |\mathbf{v}^{(w_q)}[m, k]|}{\sum_{m \in \mathcal{V}_n, k} |\tilde{\mathbf{s}}^{(w_q)}[m, k]|}, \quad (16)$$

where \mathcal{V}_n is an interval centered at n , whose length is determined based on the window size in w . Using this analysis, we define a novel TF representation V_{adap} , by selecting the mode component corresponding to the optimal window \hat{q}_n for each n :

$$V_{\text{adap}}[n, k] = \mathbf{u}^{(w_{\hat{q}_n})}[n, k], \quad \forall k. \quad (17)$$

This new TF representation is expected to be less affected by interference than a single window spectrogram, enabling an improved estimation of the IFs of the modes in an MCS when they do not cross. To achieve this, RD is performed using the classical approach introduced in [17]. As an illustration, we consider the two-chirp signal involved in Fig. 3a and evaluate the spectrogram for different window lengths $w = 15, 20, \dots, 75$ at an input SNR of 10 dB. RD is then performed on V_{adap} . In Fig. 4a, we display the image V_{adap} along with the detected ridges, demonstrating how knowledge of the interference structure enables better window size selection, facilitating more accurate IF estimation. In Fig. 4b, we illustrate the process of window length selection for two different characteristic times.

V. CONCLUSION AND PERSPECTIVES

In this work, we introduced two novel approaches for separating mode components from interference in time-frequency representations, leveraging both variational methods and deep learning. We compared a TV-based algorithm with a U-Net model trained on synthetic data, demonstrating the ability of both methods to extract the oscillatory interference while preserving the geometric structures of the modes. Building on mode-interference separation, we proposed an adaptive window selection strategy that minimizes the relative contribution of interference to construct an enhanced time-frequency representation, facilitating more accurate IF estimation.

To the best of our knowledge, applying cartoon-texture decomposition algorithms to disentangle modes from interference is a novel and original approach that warrants further investigation. Future work will focus on refining the variational framework by designing function spaces and norms better suited to the smooth, non-cartoon nature of geometric modes and the specific oscillatory patterns to be captured. Additionally, the specific noise characteristics of the spectrogram could be explicitly modeled as a third component to be separated, rather than being merged with the texture term in the current approach. Regarding the deep learning approach, alternative architectures and Plug&Play methods will be explored to enhance separation quality and mitigate irregularities.

REFERENCES

- [1] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Sig. Proc.*, vol. 51, no. 1, pp. 101–111, 2003.
- [2] C. L. Herry, M. Frasch, A. J. Seely, and H.-T. Wu, "Heart beat classification from single-lead ECG using the synchrosqueezing transform," *Physiological Measurement*, vol. 38, no. 2, pp. 171–187, 2017.
- [3] Y.-Y. Lin, H.-T. Wu, C.-A. Hsu, P.-C. Huang, Y.-H. Huang, and Y.-L. Lo, "Sleep apnea detection based on thoracic and abdominal movement signals of wearable piezoelectric bands," *IEEE J. biomedical and health informatics*, vol. 21, no. 6, pp. 1533–1545, 2017.
- [4] N. Delprat, "Global frequency modulation laws extraction from the gabor transform of a signal: A first study of the interacting components case," *IEEE Trans. speech audio proc.*, vol. 5, no. 1, pp. 64–71, 1997.
- [5] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, "Time-frequency reassignment and synchrosqueezing: An overview," *IEEE Sig. Proc. Magazine*, vol. 30, no. 6, pp. 32–41, 2013.
- [6] T. Oberlin, S. Meignen, and V. Perrier, "Second-order synchrosqueezing transform or invertible reassignment? towards ideal time-frequency representations," *IEEE Trans. Sig. Proc.*, vol. 63, no. 5, pp. 1335–1344, 2015.
- [7] S. Meignen, N. Laurent, and T. Oberlin, "One or two ridges? an exact mode separation condition for the Gabor transform," *IEEE Sig. Proc. Letters*, vol. 29, pp. 2507–2511, 2022.
- [8] V. Bruni, R. Campagna, and D. Vitulano, "Multicomponent signals interference detection exploiting HP-splines frequency parameter," *Applied Numerical Mathematics*, vol. 209, pp. 20–37, 2025.
- [9] V. Bruni, D. Vitulano, and S. Marconi, "A supervised approach for the detection of am-fm signals' interference regions in spectrogram images," *Image and Vision Computing*, vol. 138, p. 104812, Oct. 2023.
- [10] S. Zhang, M. S. R. Pavel, and Y. D. Zhang, "Crossterm-free time-frequency representation exploiting deep convolutional neural network," *Sig. Proc.*, vol. 192, p. 108372.
- [11] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition—modeling, algorithms, and parameter selection," *International J. computer vision*, vol. 67, pp. 111–136, 2006.
- [12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [13] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "Image Decomposition into a Bounded Variation Component and an Oscillating Component," vol. 22, no. 1, pp. 71–88.
- [14] J.-F. Aujol and A. Chambolle, "Dual norms and image decomposition models," *International J. computer vision*, vol. 63, pp. 85–104, 2005.
- [15] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. imag. and vis.*, vol. 20, pp. 89–97, 2004.
- [16] K. Polissano, S. Meignen, N. Laurent, and H. Leterme, "Code implementation of the present paper," 2025, <https://gricad-gitlab.univ-grenoble-alpes.fr/polisank/disentangling-modes-and-interference-in-the-spectrogram-of-multicomponent-signals>.
- [17] R. A. Carmona, W. L. Hwang, and B. Torrèsani, "Multiridge detection and time-frequency reconstruction," *IEEE Trans. Sig. Proc.*, vol. 47, no. 2, pp. 480–492, 1999.