A novel approach based on Voronoï cells to classify spectrogram zeros of multicomponent signals

Nils Laurent, Sylvain Meignen, Marcelo Alejandro Colominas, Juan M. Miramont, Francois Auger

A NOVEL APPROACH BASED ON VORONOÏ CELLS TO CLASSIFY SPECTROGRAM ZEROS OF MULTICOMPONENT SIGNALS

N. Laurent[a], S. Meignen [a], M. A. Colominas [b], J. M. Miramont [c] and F. Auger [c]

[a] LJK, University of Grenoble-Alpes and CNRS UMR 5224, F38401 Grenoble, France
[b] Institute for Research and Development in Bioengineering and Bioinformatics (IBB, UNER - CONICET), and Faculty of Engineering (UNER) Oro Verde, Entre Rios, Argentina
[c] Nantes Université, Institut de Recherche en Énergie Électrique de Nantes Atlantique (IREENA, UR 4642), F-44600 Saint-Nazaire, France

ABSTRACT

In this paper, we propose a novel approach to classify the spectrogram zeros (SZs) of multicomponent signals based on the analysis of the Voronoï cells associated with these zeros. More precisely, the characterization of the distribution of the spectrogram maxima of a complex white Gaussian noise along the edges of the Voronoï cells associated with SZs enables us to derive an algorithm to classify the different types of zeros present in the spectrogram of a multicomponent signal. Numerical applications on simulated signals confirm the relevance of the proposed classification algorithm, and an illustration on a real signal concludes the paper.

Index Terms— Time-frequency analysis, short-time Fourier transform, spectrogram zeros, Voronoï cells.

1. INTRODUCTION

The analysis of non-stationary multicomponent signals from their spectrogram has been the subject of intense research in the last two decades, mainly because the modes of these signals make up ridges in the *time-frequency* (TF) plane that are used for *instantaneous frequency* (IF) estimation [1] and mode reconstruction [2]. These ridges correspond to local maxima of the spectrogram along the frequency axis, and, for this reason, it is often acknowledged that the important information of the spectrogram is located where these maximal coefficients stand.

However, in [3], it is explained, by means of the Bargmann factorization [4, Section 3.4], that the *short-time Fourier transform* (STFT) can be viewed as an analytic function which is completely characterized by the location of its zeros [3]. Furthermore, the location of spectrogram zeros (SZs) of a *complex white Gaussian noise* (CWGN) is shown to exhibit a very specific distribution, which was used to discriminate the signal from the noise using Delaunay triangulation of SZs [3]. A deeper mathematical description of SZs of CWGNs was then provided in [5], showing that their distribution is homogeneous in the TF plane [5], meaning that, if a region is devoid of zeros, it is most probably associated with signal information. It is worth noting here that, much before these findings, SZs had already been used in TF analysis of the auditory system [6], where, using a reassignment technique, the repulsive effect of SZs was illustrated. Finally, we should also note that the analysis of the behavior of the phase of the STFT in the vicinity of SZs was investigated in [7].

In all the existing approaches to separate the noise from the signal, one either uses SZs or local maxima of the spectrogram, but not both at the same time. In the present paper, we propose an algorithm to classify SZs of an MCS in three different classes, based on the maxima of the spectrogram coefficients along the edges of the *Voronoï cells associated with SZs* (VSZs). The rationale for using such a strategy is the idea that, as remarked in [8], the largest spectrogram coefficients are located close to the edges of VSZs. The paper is organized as follows. In Sec. 2, we define the notation and basic definitions to be used throughout the paper. Then, in Sec. 3, we investigate the distribution of the maxima of the spectrogram coefficients of a CWGN along the edges of VSZs, which is the basis for the SZs classification algorithm detailed in Sec. 4. Finally, Sec. 5 is devoted to the potential use of this classification algorithm to determine how separable the modes are in the TF plane. An illustration of this last aspect on a real signal concludes the paper.

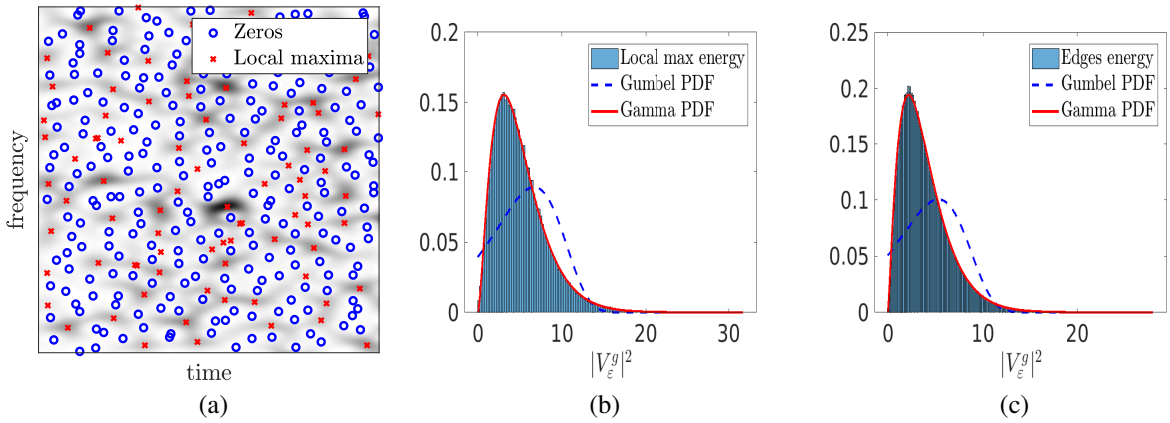
2. DEFINITIONS AND NOTATION

For $f, g \in l^1(\mathbb{Z})$, the discrete-time STFT is defined by,

$$V_f^g[n,k] = \sum_{l\in\mathbb{Z}} f[l+n]g[l]e^{-2i\pi\frac{kl}{N}}, \qquad (1)$$

in which N is the number of frequency bins. In the sequel, the type of signals we study are *multicomponent signals* (MCSs),

This work is supported by the ANR ASCETE project (France) with grant number ANR-19- CE48-0001-01.



(a) (b) (c)

Fig. 1. (a): Local maxima and zeros of the spectrogram of a CWGN; (b): Histogram of the local maxima of the coefficients of the normalized spectrogram of CWGN, plus the maximum likelihood estimate obtained assuming either a Gumbel or a Gamma prior; (c): Same as (b) but for the histogram of the maxima along the edges of VSZs of normalized spectrogram coefficients of CWGN. (d): classification error as a function of the noise level for the three signals of Fig. 2

defined by the superimposition of P modes, as follows:

$$f[n] = \sum_{p=1}^{P} f_p[n], \tag{2}$$

where $f_p[n] = a_p[n]e^{2i\pi\phi_p[n]}$, in which a_p and ϕ_p are respectively the instantaneous amplitude and instantaneous phase.

3. ON THE DISTRIBUTION OF LOCAL MAXIMA OF THE SPECTROGRAM OF A CWGN

Assume that ε is a circular CWGN with variance σ_ε^2 for its real and imaginary parts, it is well known that the coefficients of its normalized spectrogram follow a chi-squared distribution [9]:

$$\frac{|V_\varepsilon^g[n,k]|^2}{\sigma_\varepsilon^2\|g\|^2} \sim \chi_2^2. \tag{3}$$

Regarding the distribution of the local maxima of the normalized spectrogram, it was hypothesized in [3] that they should also follow a Gumbel distribution. However, to the best of our knowledge, no theoretical results are available to support this claim, and we now show numerically that a Gamma distribution is much more adapted to model the distribution of the local maxima of the normalized spectrogram than a Gumbel distribution. Indeed, in Fig.1 (a), we display the location of the SZs and of the local maxima of the spectrogram of a CWGN, and, in Fig.1 (b), the histogram of the local maxima of the normalized spectrogram built using 300 noise realizations, along with its maximum likelihood estimator (MLE) assuming a Gumbel or a Gamma prior. This confirms that the latter is more appropriate than the former to represent the distribution of local maxima.

We now would like to investigate the behavior of the maxima of normalized spectrogram along VSZ edges. In this regard, it is reported in [10] that spectrogram coefficients of a CWGN are mostly reassigned on the edges of VSZs, meaning an edge point is close to the location of a local maximum of the spectrogram in the direction given by the reassignment vector. For that reason, the maxima of the normalized spectrogram along such an edge should follow a distribution close to that of the local maxima, namely a Gamma distribution, since they somehow correspond to local maxima of the spectrogram in two different directions. More formally, let us consider a CWGN ε defined as in the previous section, z_i a zero of its spectrogram, C_ε^i its Voronoï cell. The latter is a polygon composed of J_i segments $C_\varepsilon^{i,j}$ for $j \in [\![1, J_i]\!]$, and the variable

$$\mathrm{Max}_\varepsilon^{i,j} = \max_{[n,k]\in C_\varepsilon^{i,j}} \frac{|V_\varepsilon^g[n,k]|^2}{\gamma^2}, \tag{4}$$

in which $\gamma := \sigma_\varepsilon\|g\|_2$, happens to follow the Gamma distribution $\Gamma(k,\theta)$ with $k \approx 2.28$ and $\theta \approx 1.7$, computed from MLE, the results being shown in Fig. 1 (c). From this analysis, one defines a threshold R, such that if $\mathrm{Max}_\varepsilon^{i,j} > R$ then it is not en edge in noise with probability $1 - 10^{-3}$. In the following section, we are going to use the variable defined in (4) and the threshold R to classify SZs of noisy MCSs.

4. CLASSIFICATION OF SPECTROGRAM ZEROS BASED ON VORONOÏ CELLS

4.1. Classification Algorithm

When considering a noisy MCS, SZs result from three different types of interference, either between Gaussian logons in noise (assuming the analysis window is Gaussian), between a Gaussian logon and the STFT of a component of the signal, or between the STFTs of two different signal components. In what follows, we denote these three types of SZs by noise-noise (NN), signal-noise (SN), or signal-signal (SS).

We now detail our novel algorithm to classify SZs in the three mentioned classes. For that purpose, let us consider a noisy version of f, denoted by $\tilde{f} = f + \varepsilon$, in which ε is a CWGN. First of all, we would like to compute an estimate of the normalized spectrogram and thus $\gamma := \sigma_\varepsilon \|g\|_2$, for which the robust estimator [11, 12], $\widehat{\gamma} := \text{median}_{[n,k]} \frac{\left|\Re\left\{V_{\tilde{f}}^g[n,k]\right\}\right|}{0.6745}$, is often used. However, such an estimator is biased, since some coefficients corresponding to the signal are taken into account in the estimation. So to improve the latter, we propose to remove the coefficients that most probably correspond to the signal, by first defining $\mathcal{S}_{\tilde{f}} = \{[n,k];\ |V_{\tilde{f}}^g[n,k]| < 3\widehat{\gamma}\}$ as, from the χ_2^2 table, coefficients $|V_{\tilde{f}}^g[n,k]| > 3\gamma$ are associated with signal with probability larger than 0.99, and then by considering the improved estimate for γ given by:

$$\tilde{\gamma} := \underset{[n,k]\in\mathcal{S}_{\tilde{f}}}{\text{median}} \frac{\left|\Re\left\{V_{\tilde{f}}^g[n,k]\right\}\right|}{0.6745}. \tag{5}$$

In the sequel, a normalized spectrogram coefficient corresponds to $\frac{|V_{\tilde{f}}^g[n,k]|^2}{\tilde{\gamma}^2}$ for some $[n,k]$, and using the same notation as for Voronoï cells in the pure noise situation, we define the local maxima along $C_{\tilde{f}}^{i,j}$ as:

$$\text{Max}_{\tilde{f}}^{i,j} = \max_{[n,k]\in C_{\tilde{f}}^{i,j}} \frac{|V_{\tilde{f}}^g[n,k]|^2}{\tilde{\gamma}^2}, \tag{6}$$

and then propose the following algorithm to classify SZs in one of the three classes introduced in the beginning of this section:

Algorithm 1 Classification of Spectrogram zeros

Input: $\mathcal{Z}$ set of spectrogram zeros, threshold R
for $z_i \in \mathcal{Z}$ **do**
  $C_{\tilde{f}}^i$ Voronoï cell, with J_i edges, associated with z_i
  $C = \text{card}\,\{j \in [\![1, J_i]\!], \text{Max}_{\tilde{f}}^{i,j} > R\}$
  if $C = J_i$ **then**
    $z_i \in SS$
  else
    if $3 \leq C$ **then**
      $z_i \in SN$
    else
      $z_i \in NN$
    end if
  end if
end for

Indeed, when a SZ is of SS type, it is most of the time associated with a specific TF pattern called *time-frequency bubble* [13]. Such a structure corresponds to a SZ circled by normalized spectrogram coefficients with high amplitude, and the edges of the corresponding VSZ follow, more or less, these coefficients in the TF plane. Thus, a normalized spectrogram maximum along an edge of such a VSZ should be above R. When a SZ is of type SN, one edge of its Voronoï cell follows the ridge of the mode of the signal involved in the interference, implying that at least the normalized spectrogram maxima along three different edges should be above R. Finally, to improve the efficiency of the proposed algorithm, a SZ of type SS is changed into SN when there is a SZ of NN type in a neigbhoring cell.

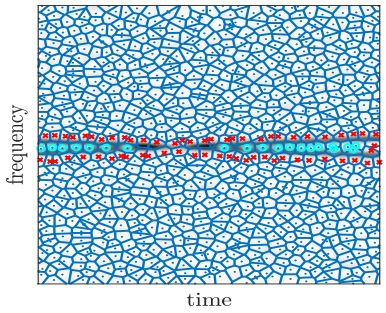


Fig. 2. Voronoï cells associated with SZs of two crossing modes, the dots, the stars and the circle correspond to SZs of NN, SN, and SS types respectively

An illustration of the behavior of the algorithm is given in Fig.2, for the signal made of two crossing linear chirps. We remark that the classification of the SZs in three different types seems correct, but we need to quantify classification errors more precisely, which is done in the next section.

4.2. Validation on Synthetic MCSs

In this section, we propose to validate the previous algorithm by checking that a SZ is correctly classified into one of the three classes. To define what should be the class of a SZ, we associate with the mode f_p the region $\mathcal{H}_{f_p} := \{[n,k]; |V_{f_p}^g[n,k]| > 3\gamma\}$. Then, for each SZ z_i, we count the number of these regions that are intersected by VSZ $C_{\tilde{f}}^i$, namely:

$$D^i := \text{card}\{p \in [\![1, P]\!]; C_{\tilde{f}}^i \cap \mathcal{H}_{f_p} \neq \emptyset\}. \tag{7}$$

The three perfect classes are then defined by

$$\begin{aligned} \mathcal{Z}_{SS} &:= \{z_i,\ D^i > 1\}, \\ \mathcal{Z}_{SN} &:= \{z_i,\ D^i = 1\}, \\ \mathcal{Z}_{NN} &:= \{z_i,\ D^i = 0\}. \end{aligned} \tag{8}$$

Then, a zero z_i of type T_i is said to be correctly classified, if $z_i \in \mathcal{Z}_{T_i}$, and we thus measure the classification error by computing $1 - \frac{1}{\text{card}(\mathcal{Z})} \sum_i \mathbb{1}_{\{\mathcal{Z}_{T_i}\}}(z_i)$.

To evaluate the quality of our classification method, when the input SNR varies, we depict in Fig. 3 (d), the classification error in percentage associated with the signals of Fig. 3

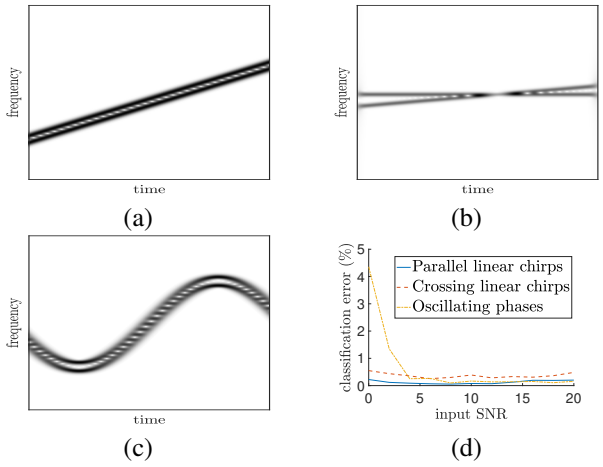


Fig. 3. (a): Spectrogram of two close parallel linear chirps; (b): Spectrogram of two crossing linear chirps; (c): Spectrogram of two close oscillating modes; (d): classification results

(a) to (c). The proposed method performs best for the two parallel linear chirps, while for crossing linear chirps, the error is slightly higher, but in both cases it remains stable when the noise level varies. Finally, for the signal with oscillating phases, the error is, for most input SNRs, lesser than 1%, but increases when the input SNR decreases below 5 dB. This is due to wrong classifications in the regions where the frequency variations of the modes are strong, which are known to be more sensitive to noise [14]. In such cases, SZs of SS type may be classified in the SN class.

5. NUMERICAL EXPERIMENTS

5.1. Investigating Mode Separability using SZs classification

Our goal in this section is to show that the classification of SZs into three different classes is a good indicator of how well the mode are separated in the TF plane. Considering the signals of Fig. 3 (a) and (b), for the parallel linear chirps we move the modes away one from another and see how the classification results varies, with respect to the mode distance in the TF plane measured by $d = \phi_2'[n] - \phi_1'[n]$, constant in that cases, and noise level. Regarding the signal of Fig. 3 (b), corresponding to a pure harmonic and a linear chirp, we make the same computation as previously with respect to the chirp rate r of the latter and noise. As expected, the results depicted in Fig. 4, show that while it is possible to separate the two pure tones by increasing the distance between them, such is not the case with the signal made of crossing modes.

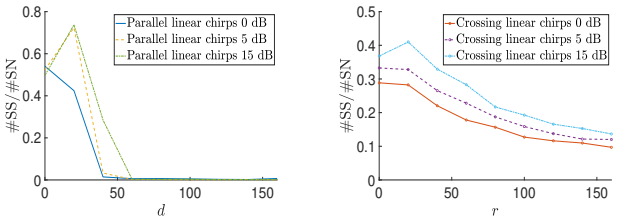


Fig. 4. Value of $\frac{\#SS}{\#SN}$ averaged over 30 realizations for the signals of Fig. 3 (a) and 3 (b), at SNRs 0 dB, 5 dB and 15dB.

5.2. Investigating Mode Separation in Bat Echolocation Signal

We here consider a bat echolocation signal containing 256 samples corresponding to a time span of 1,6 ms, which allows frequency analysis up to 80 kHz. Because of a too low frequency sampling, the signal contains some aliasing. In Fig. 3 (c), the spectrogram of this signal is depicted with VSZs superimposed. Four modes with sufficiently high energy can be observed, which appear to be surrounded by SZs classified as SN or SS when the modes are close to each other, the location of SS SZs telling us where the modes will be harder to separate. Note that, in the first time instants, some zeros are classified as SS because of aliasing, corresponding to modes with frequencies above 80 kHz.

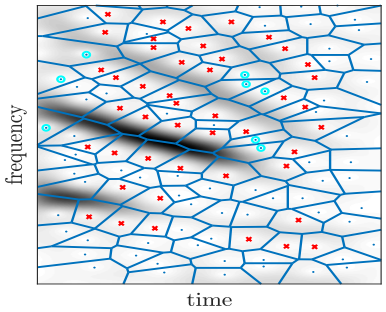


Fig. 5. VSZs associated with the bat signal, along with classification results of these zeros in the three classes.

6. CONCLUSION

In this paper, we proposed a method to classify the spectrogram zeros of noisy multicomponent signals. These are generated by interference of different types which we denoted by noise-noise, signal-noise or signal-signal. Based on a thorough study of the distribution of spectrogram coefficients on the edges of the Voronoï cells associated with spectrogram zeros in pure noise, we were able to accurately classify the spectrogram zeros of noisy multicomponent signals in one of the above-mentioned classes.

7. REFERENCES

[1] L. Stankovic, M. Dakovic, and V. Ivanovic, "Performance of spectrogram as IF estimator," *Electronics Letters*, vol. 37, no. 12, pp. 797–799, 2001.

[2] S. Meignen and D.-H. Pham, "Retrieval of the modes of multicomponent signals from downsampled short-time Fourier transform," *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6204–6215, 2018.

[3] P. Flandrin, "Time–frequency filtering based on spectrogram zeros," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2137–2141, 2015.

[4] K. Gröchenig, *Foundations of time-frequency analysis*. Springer Science & Business Media, 2001.

[5] R. Bardenet, J. Flamant, and P. Chainais, "On the zeros of the spectrogram of white noise," *Applied and Computational Harmonic Analysis*, vol. 48, no. 2, pp. 682–705, 2020.

[6] T. J. Gardner and M. O. Magnasco, "Sparse time-frequency representations," *Proceedings of the National Academy of Sciences*, vol. 103, no. 16, pp. 6094–6099, 2006.

[7] P. Balazs, D. Bayer, F. Jaillet, and P. Søndergaard, "The pole behavior of the phase derivative of the short-time Fourier transform," *Applied and Computational Harmonic Analysis*, vol. 40, no. 3, pp. 610–621, 2016.

[8] P. Flandrin, "On spectrogram local maxima," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 3979–3983.

[9] L. H. Koopmans, *The spectral analysis of time series*. Elsevier, 1995.

[10] P. Flandrin, "The sound of silence: Recovering signals from time-frequency zeros," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 544–548.

[11] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[12] L. Sachs, *Applied Statistics: A Handbook of Techniques*. New York: Springer-Verlag, 1984, vol. p.253.

[13] N. Delprat, "Global frequency modulation laws extraction from the Gabor transform of a signal: A first study of the interacting components case," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 64–71, 1997.

[14] N. Laurent and S. Meignen, "A novel ridge detector for nonstationary multicomponent signals: development and application to robust mode retrieval," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3325–3336, 2021.