



# time-frequency analysis of noisy multicomponent signals : contributions to ridge detection, mode reconstruction, chirp rate estimation, and interference localization.

Nils Laurent

## ► To cite this version:

Nils Laurent. time-frequency analysis of noisy multicomponent signals : contributions to ridge detection, mode reconstruction, chirp rate estimation, and interference localization.. General Mathematics [math.GM]. Université Grenoble Alpes [2020-..], 2022. English. NNT : 2022GRALM060 . tel-04119354

**HAL Id: tel-04119354**

**<https://theses.hal.science/tel-04119354>**

Submitted on 6 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : MSTII - Mathématiques, Sciences et technologies de l'information, Informatique

Spécialité : Mathématiques Appliquées

Unité de recherche : Laboratoire Jean Kuntzmann

### Analyse temps-fréquence de signaux multicomposantes bruités :

contribution à la détection de ligne de crête, reconstruction des modes, estimation de la variation de fréquence et localisation de l'interférence.

### Time-frequency analysis of noisy multicomponent signals:

contributions to ridge detection, mode reconstruction, chirp rate estimation, and interference localization.

Présentée par :

**Nils LAURENT**

#### Direction de thèse :

**Sylvain MEIGNEN**

Université Grenoble Alpes

Directeur de thèse

**Bertrand RIVET**

Université Grenoble Alpes

Co-directeur de thèse

**Julie FONTECAVE-JALLON**

Université Grenoble Alpes

Co-encadrante de thèse

#### Rapporteurs :

**MARIA SANDSTEN**

Professeur, Lunds Universitet

**ROLAND BADEAU**

Professeur, TELECOM PARIS

Thèse soutenue publiquement le **29 septembre 2022**, devant le jury composé de :

**PIERRE CHAINAIS**

Professeur, ECOLE CENTRALE LILLE

Président

**MARIA SANDSTEN**

Professeur, Lunds Universitet

Rapporteuse

**ROLAND BADEAU**

Professeur, TELECOM PARIS

Rapporteur

**PATRICK FLANDRIN**

Directeur de recherche, CNRS DELEGATION RHONE AUVERGNE

Examineur

**JÉRÔME MARS**

Professeur des Universités, GRENOBLE INP

Examineur

**SYLVAIN MEIGNEN**

Maître de conférences HDR, GRENOBLE INP

Directeur de thèse

**BERTRAND RIVET**

Maître de conférences HDR, GRENOBLE INP

Co-directeur de thèse

#### Invités :

**Julie FONTECAVE-JALLON**

Maître de conférences, UNIVERSITE GRENOBLE ALPES





# Acknowledgment

I would first like to express my warmest thanks to my supervisor, Sylvain Meignen, who was inspiring, dedicated and benevolent both as a teacher and as a supervisor. I would also like to thank my co-supervisors, Bertrand Rivet and Julie Fontecave-Jallon, for their availability, kindness and for the helpful discussions.

I thank Maria Sandsten and Roland Badeau, for their time reporting my thesis. I further thank them with the other members of the jury; Patrick Flandrin, Jérôme Mars and Pierre Chainais for rich interaction and their precious advices.

I thank my colleagues in the Laboratoire Jean Kuntzmann and my collaborators, for the very rich and friendly discussions. I also thank them for the advices they gave me for teachings. I also thank those who encouraged me in my studies, including some of my previous teachers and dear friends. Finally, I thank my family for all the trust and freedom they gave me.





# Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Short-Time Fourier Transform and Spectrogram . . . . .	1
1.2. Reassignment and Synchrosqueezing . . . . .	3
1.3. Multicomponent signals . . . . .	5
1.4. Outline of the Following Chapters . . . . .	8
1.5. List of published articles linked to the contributions of this thesis . . . . .	9
<b>2. Mode Retrieval Based on Robust Ridge Detection and Linear Chirp Approximation</b>	<b>11</b>
2.1. Introduction . . . . .	11
2.2. Basics of Ridge Detection . . . . .	12
2.3. Definition of a New Robust Ridge Detector . . . . .	15
2.4. Linear Chirp Approximation . . . . .	23
2.5. Approximation of $\phi'$ and $\phi''$ Based on FSST2 and Ridge Detection . . . . .	24
2.6. Linear Chirp Based Mode Retrieval Technique . . . . .	25
2.7. Alternative Techniques for Mode Retrieval: HT, SSR-HT . . . . .	26
2.8. Numerical Applications . . . . .	28
2.9. Application to Gravitational-Wave Signals . . . . .	31
2.10. Conclusion . . . . .	33
<b>3. Time Frequency Analysis of Interference</b>	<b>35</b>
3.1. Introduction . . . . .	35
3.2. On the Existence of Time Frequency Bubbles . . . . .	35
3.3. Localizing TFBs Based on a New RD . . . . .	40
3.4. Results . . . . .	44
3.5. Analysis of Interference Based on the Zeros of the Spectrogram . . . . .	49
3.6. Perspectives for a Zero Based Technique to Identify Interference . . . . .	50
3.7. Conclusion . . . . .	52

<b>4. Towards Adaptive and Robust Synchrosqueezing</b>	<b>55</b>
4.1. Introduction . . . . .	55
4.2. Definition of FSSTN . . . . .	55
4.3. Adaptive selection of the order . . . . .	58
4.4. Conclusion . . . . .	63
<b>5. Study of the Chirp Rate Estimate in Case of Noise</b>	<b>65</b>
5.1. Evaluating the CR Filtering Process . . . . .	71
5.2. Frequency Variation of a Voice Signals . . . . .	75
5.3. Conclusion . . . . .	78
<b>6. SST Based Heart Rate Estimation</b>	<b>79</b>
6.1. Introduction . . . . .	79
6.2. Sparsity of Harmonic Signals and Application to ECG Signals . . . . .	80
6.3. Conclusion . . . . .	86
<b>7. Conclusion</b>	<b>89</b>
7.1. Contributions . . . . .	89
7.2. Future research . . . . .	89
<b>Bibliography</b>	<b>91</b>
<b>Appendices</b>	<b>101</b>
<b>A. Time Frequency Analysis of Interference</b>	<b>I</b>
A.1. Proof of Proposition 3.2.1 . . . . .	I
A.2. Proof of Proposition 3.2.2 . . . . .	III
<b>B. Towards Robust and Adaptive Synchrosqueezing</b>	<b>IX</b>
B.1. Proof of Proposition 5.0.1 . . . . .	IX
B.2. Proof of Proposition 5.0.2 . . . . .	X
B.3. Proof of Proposition 5.0.3 . . . . .	XI

## List of Acronyms

<b>TFR</b>	<i>time frequency representation</i>
<b>TF</b>	<i>time frequency</i>
<b>FT</b>	<i>Fourier transform</i>
<b>STFT</b>	<i>short-time Fourier transform</i>
<b>RM</b>	<i>reassignment method</i>
<b>SST</b>	<i>synchrosqueezing transform</i>
<b>FSST</b>	<i>Fourier-based synchrosqueezing transform</i>
<b>FSST2</b>	<i>second order Fourier-based synchrosqueezing transform</i>
<b>FSST4</b>	<i>fourth-order Fourier-based synchrosqueezing transform</i>
<b>FSSTN</b>	<i>high order Fourier-based synchrosqueezing transform</i>
<b>MCS</b>	<i>multicomponent signal</i>
<b>IF</b>	<i>instantaneous frequency</i>
<b>CR</b>	<i>chirp rate</i>
<b>ITF</b>	<i>ideal time frequency representation</i>
<b>MR</b>	<i>mode retrieval</i>

## Acronyms of Chapter 2

<b>RD</b>	<i>ridge detection</i>
<b>LMMF</b>	<i>local maximum along the frequency axis</i>
<b>S-RD</b>	<i>simple ridge detector</i>
<b>MB-RD</b>	<i>modulation based ridge detector</i>
<b>RRP</b>	<i>relevant ridge portion</i>
<b>RP</b>	<i>ridge portion</i>

**RRP-RD** *relevant ridge portion based ridge detector*

**SNR** *signal to noise ratio*

**BA** *basin of attraction*

**LCR** *linear chirp based retrieval*

**HT** *hard thresholding*

**SSR-HT** *shifted-symmetrized-regularized hard-thresholding*

**S-LCR-MR** *MR based on S-RD and LCR*

**RRP-LCR-MR** *MR based on RRP-RD and LCR*

**RRP-MR** *MR based on RRP-RD and HT*

**FSST4-MR** *MR based on FSST4*

### **Acronyms of Chapter 3**

**LMF** *local maximum of the spectrogram along the frequency axis*

**TFB** *time frequency bubble*

**LEF** *local extremum of the spectrogram along the frequency axis*

**R-RD** *redundant ridge detector*

**R-RDN** *redundant ridge detector in noise*

**GT** *ground truth*

**PH** *pure harmonic*

### **Acronyms of Chapter 4**

**PSD** *power spectral density*

### **Acronyms of Chapter 6**

**HR** *heart rate*

**FF** *fundamental frequency*

**ECG** *electrocardiogram*

**NMF** *non-negative matrix factorization*

**EMD** *earth mover distance*

**bpm** *beats per minute*



# List of Figures

1.1.	(a) STFT of $f(t) = \exp(2i\pi(50(t^4 + t^3 + t^2) + 100t))$ ; (b) instantaneous chirp rate $CR(t) = 600t^2 + 300t + 100$ . . . . .	5
1.2.	Spectrogram of a bat echolocation signal. . . . .	6
1.3.	(a): STFT of a MCS, (b) ITF of the signal of (a). . . . .	7
2.1.	LMMFs corresponding to the three largest STFT modulus maxima for each time instant along with the true IF (dashed line) of the linear chirp (SNR= -10 dB ) . . . . .	14
2.2.	(a): from top to bottom: STFT of a linear chirp, of a signal with cosine phase, and of a signal with oscillatory phase; (b): $\mathcal{P}(\beta)$ for these three signals when the input SNR equals -10 dB as well as the proportion of time indices $n$ , at which the relevant LMMF with ordinate in $\mathcal{I}[n]$ corresponds to the largest STFT modulus maximum (curves denoted by "Max, linear chirp", "Max, cosine", and "Max, modulated cosine"). The results are averaged over 100 noise realizations. . . . .	18
2.3.	(a): STFT modulus of two noisy linear chirps (SNR = -10 dB, $\sigma = 0.0188$ ); (b): STFT modulus of two noisy modes with cosine phase, with different modulation (SNR = -10 dB, $\sigma = 0.0175$ ); (c): STFT modulus of a signal made of a linear chirp and a mode with exponential phase (SNR = -10 dB, $\sigma = 0.0241$ ); (d): $(s_p^{fin})_{p=1,2}$ computed for the signal displayed in (a); (e): same as (d) but for the signal whose STFT modulus is displayed in (b); (f): same as (d) but for the signal whose STFT modulus is displayed in (c). . . . .	21



- 2.4. (a): Comparison between S-RD, MB-RD and RRP-RD, for the mode  $f_1$  of the signal of Figure 2.3 (a), computation of output SNR between IF  $\phi'_1$  and estimated IF with respect to input SNR (the results are averaged over 40 realizations of noise); (b): same as (a) but for the mode  $f_1$  of the signal of Figure 2.3 (b); (c): same as (a) but for the mode  $f_1$  of the signal of Figure 2.3 (c); (d): same as (a) but the mode  $f_2$  of the signal of Figure 2.3 (a); (e): same as (b) but the mode  $f_2$  of the signal of Figure 2.3 (b); (f): same as (c) but the mode  $f_2$  of the signal of Figure 2.3 (c). . . . . 23
- 2.5. (a): STFT of  $f(t) = e^{2i\pi(250t+568\frac{t^2}{2})}$ ; (b): STFT of  $f(t) = e^{2i\pi(1200t+60\cos(2\pi t))}$ ; (c): STFT of  $f(t) = e^{2i\pi(2000t+238\cos(2\pi t))}$ . . . . . 29
- 2.6. (a): Signal reconstruction results for the signal of Figure 2.5 (a) using HT, SSR-HT and LCR on the ridge provided by S-RD. the mode is reconstructed using either (1.7, page 3) or (1.8, page 3) ; (b): Same as (a) but for the signal depicted in Figure 2.5 (b); (c): Same as (a) but for the signal depicted in Figure 2.5 (c). . . . . 29
- 2.7. (a): For each mode  $p = 1, 2$ , output SNR between mode  $f_p$  of signal of Figure 2.3 (a) and reconstructed mode using either S-LCR-MR or RRP-LCR-MR (the results are averaged over 40 noise realizations); (b): same but with signal of Figure 2.3 (b); (c): same but with signal of Figure 2.3 (c); . . . . . 30
- 2.8. (a): computation of the output SNR associated with the estimation of  $\phi'$  with  $s^{fin}$  (computed for different values of  $\lambda$ ) or with  $\hat{\omega}^{[2]}$  for the first mode of Figure 2.3 (a); (b): computation of the SNR associated with the estimation of  $\phi''$  with  $(s^{fin})'$  (computed for different values of  $\lambda$ ) or with  $\hat{q}_{\tilde{f}}$  for the first mode of Figure 2.3 (a). . . . . 30
- 2.9. (a): STFT of signal made of two close chirps; (b): Reconstruction results using (1.7, page 3) for each of the two modes depicted in (a) using S-RD. 31
- 2.10. (a): STFT modulus ( $\sigma = 0.0105$ ) of the Hanford signal along with the ridge obtained with RRP-RD and FSST4; (b): illustration of signal reconstruction based on RRP-LCR-MR and the numerical relativity; (c): SNR corresponding to the reconstruction of the signal using either RRP-LCR-MR, RRP-MR or FSST4-MR (the ground truth is assumed to be the one produced by numerical relativity). The results are averaged over 40 realizations of noise. . . . . 32

3.1. (a): spectrogram of two parallel pure harmonic modes with the same amplitude, when there exists three LEFs at each time instant (the ridges corresponding to the chaining of LMFs are also superimposed); (b): spectrogram of two parallel pure harmonic modes exhibiting TFBs; (c): same as (a) but when the pure harmonic modes have different amplitude (three LEFs are still present at each time instant but the symmetry no longer exists). In each case, we also plot the zeros and the local maxima of the spectrogram. . . . .	38
3.2. (a): Spectrogram of two parallel linear chirps with the same amplitude, when there always exists three extrema in the direction $D$ ; (b): Spectrogram of two parallel linear chirps when TFBs are present. Three extrema no longer exist in the direction $D$ ; (c): same as (a) but when the linear chirps have different amplitude (three extrema are still present along the direction $D$ ). In (a) to (c), we also plot the zeros and the local maxima of the spectrogram. . . . .	40
3.3. (a): ridges detected using S-RD; (b): ridges detected using R-RD. . . . .	42
3.4. (a): STFT of a three mode signal along with the ridges superimposed, in the definition of the Gaussian window $\sigma = 0.02$ , the stars corresponds to the locations of the merging of ridges (R-RD is used for ridge detection); (b): same as (a) but with $\sigma = 0.03$ ; (c): same as (a) but with $\sigma = 0.04$ ;(d): study of the same signal as in (a) with some white Gaussian added noise (input SNR 20 dB), and $\sigma = 0.04$ in the definition of the Gaussian window; (e): same as (d) except the input SNR equals 10 dB;(f): same as (d) except the input SNR equals 5 dB. . . . .	44
3.5. (a): TFB detection for the signal of Figure 3.1 (a) either in the noiseless case or for different noise levels, using the function $R$ , along with the ground truth; (b): TFB detection for the signal of Figure 3.1 (c) using either $R$ or the ground truth. The results are averaged over 10 realizations of the noise. . . . .	46

3.6.	(a): Interference mask associated with the constant and linear chirps of Figure 3.4, denoted by zone 1; (b): Interference mask associated with the linear chirp and the mode with sinusoidal phase of Figure 3.4, denoted by zone 2; (c): top graph: percentage of detection of TFBs in zone 1, zone 2, and elsewhere (false detections) when the noise level varies, bottom graph: percentage of points associated with the detection of TFBs in the different zones when the noise level varies (the number of noise realization equals 30). . . . .	47
3.7.	(a): STFT of a voice signal (recording 62 of the Saarbrücken Voice Database [65]) along with the ridges associated with 95% of the total energy (no TFB is detected); (b): same as (a) but when 98% of the total energy is located in the vicinity of the ridges (a TFB corresponding to the white star symbol is detected); (c): same as (a) but in the presence of noise (a TFB is detected is detect in the same TF region as in (b)). . . . .	48
3.8.	Illustration of the zero based Delaunay triangulation, where the red circles represent the zeros of the spectrogram. . . . .	50
3.9.	(a): Magnitude of the STFT of two PHs affected by noise, with superimposed Voronoi cells; (b): Zoom on (a) on a Voronoi cell associated with noise; (c): Zoom on (a) on a Voronoi cell associated with interference; (d): Zoom on (a) on a Voronoi cell associated with signal. . . . .	51
3.10.	Histogram associated with the selection of the maximum value on $\approx 80$ 000 segments of Voronoi cells in the case of pure noise. . . . .	52
4.1.	(a): magnitude of the STFT of a mode with an oscillating phase; (b): $E_N$ for $2 \leq N \leq 4$ for the signal whose STFT is depicted in (a) . . . . .	60
4.2.	(a): STFT magnitude of a mode with dumped sinusoidal frequency and amplitude $A = 1$ ; (b),(c),(d): $E_2, E_3$ and $E_4$ for the mode corresponding to (a), respectively. . . . .	61
4.3.	(a): output SNR corresponding to $\text{SNR}(f, R_f^N - f)$ for $N = 2, \dots, 4$ or to $\text{SNR}(f, R_f^N - f)$ for the signal of Figure 4.1 (a); (b): same as (a) but for the signal of Figure 4.2 (a). . . . .	62
4.4.	(a): modulus of the STFT of the sum of two modes with varying amplitude; (b): mode reconstruction results as a function of input SNR. . . . .	62
4.5.	(a): STFT modulus of a bat echolocation call; (b): SNR corresponding to signal reconstruction from the coefficients on the ridges of the 3 most energetic detected modes. . . . .	63

5.1.	(a): STFT modulus of a noisy linear chirp with the STFT ridge superimposed; (b): same as (a) but for a noisy chirp with fourth order polynomial phase; (c): same as (a) but for a noisy mode with oscillating phase; (d): CR estimator $\hat{q}_{f+\varepsilon}$ or $\bar{q}_{f+\varepsilon}$ estimated on the STFT ridge of the noisy signal whose STFT modulus is represented in (a), along with the true CR, for $t$ between 0.3 and 0.5; (e): same as (d) but for the signal whose STFT modulus is represented in (b); (f): same as (d) but for the signal whose STFT modulus is represented in (c). The input SNR is 10 dB. . . . .	68
5.2.	(a): boxplot corresponding to the output SNR associated with CR estimators $\hat{q}_{f+\varepsilon}$ or $F(\hat{q}_{f+\varepsilon})$ evaluated on the STFT ridges of the noisy signal corresponding to Figure 5.1 (a) over 100 realizations of the noise when the input SNR varies; (b): same as (a) but for the noisy signal corresponding to Figure 5.1 (b); (c): same as (a) but for the noisy signal corresponding to Figure 5.1 (c); (d): $\hat{q}_{f+\varepsilon}$ and $F(\hat{q}_{f+\varepsilon})$ computed for a noisy version of the signal of Figure 5.1 (a), along with the ground truth $\phi''(t)$ (input SNR = 10 dB); (e): same as (d) but for the signal of Figure 5.1 (b); (f): same as (d) but for the signal of Figure 5.1 (c). . . . .	72
5.3.	(a): magnitude of the DFT of $\phi''$ for the signals of Figure 5.1 (a) and Figure 5.1 (b); (b): same as (a) but for the signal of Figure 5.1 (c). . . . .	73
5.4.	(a): percentage of improvement for the LCR technique computed using (5.27) for the signal of Figure 5.1 (a) where the estimates for $(\phi'', \phi')$ are $(F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t))), \omega_{f+\varepsilon}^{[2]}(t, R_{f+\varepsilon}(t))), (F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t))), \phi')$ , or $(\phi'', \phi')$ ; (b): same as (a) but for the signal of Figure 5.1 (b); (c): same as (a) but for the signal of Figure 5.1 (c) . . . . .	74
5.5.	ridge corresponding to the fundamental frequency superimposed on the STFT of the recording 62 taken from [65]. . . . .	76
5.6.	(a): for 100 realizations of the noise, the range corresponding to the 5% and 95% quantiles of $\hat{q}_{f+\varepsilon}$ evaluated on the STFT ridge corresponding to the fundamental frequency of the noisy signal whose STFT modulus is displayed in Figure 5.5, $\hat{q}_f$ on the ridge corresponding to the fundamental frequency of the noiseless signal being also superimposed (input SNR = 5 dB for the noise); (b): same as (a) but with input SNR= 10 dB; (c): same as (a) but with input SNR= 15 dB. (d): same as (a) but using the filtered estimator $F(\hat{q}_{f+\varepsilon})$ instead of $\hat{q}_{f+\varepsilon}$ , $F(\hat{q}_f)$ is also superimposed ; (e): Same as (d) but with input SNR= 10 dB; (f): same as (d) but with input SNR= 15 dB. . . . .	77

6.1.	(a): STFT of an ECG generated following [92]; (b): FSST2 of the same ECG; (c): Dictionary matrix $\mathbf{W}^1$ used with STFT-based approach; (d): Dictionary matrix $\mathbf{W}^0$ used with FSST2-based approach. For the sake of legibility, in (c) and (d) we do not normalize the column of the matrices.	81
6.2.	In these figures, the HR varies around 65 bpm, which corresponds to the index 35 of the dictionary. (a): modulus of STFT, the white curve corresponds to the indices associated with the minimal EMD ( $\hat{B} = 39.61\%$ ); (b): same as (a) but for FSST ( $\hat{B} = 94.34\%$ ) (c): same as (a) but for FSST2 ( $\hat{B} = 75.4\%$ ). . . . .	85
6.3.	(a): Estimation of HR mean from $\hat{\mathbf{i}}$ . (b): Estimation of the corresponding standard deviation. . . . .	86
6.4.	(a) FSST based EMD of a real ECG signal. (b) HR detection on the same signal using $\gamma = 2.5$ . . . . .	86

# List of Tables

6.1. Computation of $\hat{B}$ on syntethic ECG signals with different HR means when either STFT, FSST or FSST2 are used to compute $\mathbf{X}$ . . . . .	84
--	----



# Notations

$\ \cdot\ _p$	$L^p$ norm, where $p \in \{1, 2\}$
$\ \cdot\ _\infty$	$L^\infty$ norm
$x \approx y$	$x$ is an approximation of $y$
$\lfloor x \rfloor$	Largest $n \in \mathbb{Z}$ such that $n \leq x$
$\lceil x \rceil$	Smallest $n \in \mathbb{Z}$ such that $n \geq x$
$\lfloor x \rfloor$	Rounding of $x$
$\text{supp}(f)$	Support of the function $f$
$\mathbb{E}\{\cdot\}$	Expectation of a random variable
$\Re(z)$	Real part of the complex number $z$
$\Im(z)$	Imaginary part of the complex number $z$
$z^*$	Complex conjugate of $z \in \mathbb{C}$
$i$	Complex number that verifies $i^2 = -1$
$\delta$	Dirac distribution

## Sets

$\mathbb{N}$	Positive integers
$\mathbb{Z}$	Integers
$\mathbb{R}$	Real numbers
$\mathbb{R}^+$	Positive real numbers
$\mathbb{C}$	Complex numbers

## Spaces

$L^1(\mathbb{R})$	Lebesgue-integrable functions
$L^2(\mathbb{R})$	Squared Lebesgue-integrable functions
$S(\mathbb{R})$	Schwartz space
$S'(\mathbb{R})$	Tempered distribution



### Operators

$Id$	Identity
$f'$	Derivative of $f$
$f^{(n)}$	Derivative of order $n$
$\partial_t f$	Partial derivative of $f$ with respect to $t$

### Transforms

$\mathcal{F}_f$	Fourier transform of $f$
$V_f^g$	Short-time Fourier transform of $f$ with a window $g$
$W_f$	Wigner-Ville distribution of $f$

# 1. Introduction

The mathematical *frequency representation* of signals was first introduced by Fourier [1], and initially used to study the propagation of heat between solids [2]. The *Fourier transform* (FT) was thus defined as a tool to perform frequency analysis, but real signals often have time varying frequencies which FT cannot capture. To analyze such signals, *time-frequency representations* were considered, and transforms such as the *short-time Fourier transform* (STFT) [3], *spectrogram* and *wavelet transform* [4] were introduced, where the latter is effective for analyzing the information content of images [5]. Such transforms however suffer from Heisenberg's uncertainty [6, 7], meaning the signal cannot be arbitrarily well localized both in time and frequency. To circumvent this problem, the *Wigner-Ville transform* was defined [8], but at the cost of interfering artifacts that hampered the readability of this representation [9]. As a consequence, other transforms came to light, as the reallocation method [10] and the synchrosqueezing transform [11], detailed later in this chapter.

## 1.1. Short-Time Fourier Transform and Spectrogram

The Fourier transform provides a mathematical representation of the frequencies of a signal. Given a signal  $f \in L^1(\mathbb{R})$ , it is defined by [12, Chapter 3],

$$\mathcal{F}_f(\eta) = \int_{\mathbb{R}} f(\tau) e^{-2i\pi\tau\eta} d\tau. \quad (1.1)$$

The frequency representation given by  $\mathcal{F}_f$  does not depend on time. To study signals with varying frequencies, other transforms are required. In this thesis, we are going to study signals with both varying frequency and amplitude, namely *non-stationary* signals. Therefore, in order to perform their *time frequency* (TF) analysis, we recall the definition of the most commonly used *time frequency representations* (TFRs).

### 1.1.1. Continuous Formalism

The modified STFT of a signal  $f \in L^2(\mathbb{R})$  using a window  $g \in L^2(\mathbb{R})$  is an invertible linear and continuous application, defined by [12, Chapter 3],

$$V_f^g(t, \eta) = \int_{\mathbb{R}} f(u)g(u-t)e^{-2i\pi\eta(u-t)}du, \quad (1.2)$$

and the *spectrogram* by  $|V_f^g(t, \eta)|^2$ . If  $f$  is analytic, i.e.  $f$  only contains positive frequencies, and if  $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ , and  $g(0) \neq 0$ , then by applying Plancherel theorem one obtains,

$$f(t) = \frac{1}{g(0)} \int_{\mathbb{R}^+} V_f^g(t, \eta) d\eta. \quad (1.3)$$

Or if  $g \neq 0$ , as  $\{u \mapsto g(t-u)e^{2i\pi(t-u)\eta}\}$  is a frame of  $L^2(\mathbb{R})$ ,

$$f(t) = \frac{1}{\|g\|_2} \int_{\mathbb{R}^2} V_f^g(u, \eta)g(t-u)e^{2i\pi(t-u)\eta}dud\eta. \quad (1.4)$$

Or if  $\mathcal{F}_f \in L^1(\mathbb{R})$ ,  $g \in L^1(\mathbb{R})$  and  $\int_{\mathbb{R}} g(\tau)d\tau \neq 0$ , using the overlap-add method [13, 14],

$$f(t) = \frac{1}{\int_{\mathbb{R}} g(\tau)d\tau} \int_{\mathbb{R}^2} V_f^g(u, \eta)e^{2i\pi(t-u)\eta}dud\eta. \quad (1.5)$$

### 1.1.2. Discrete Formalism

In order to proceed with the practical study of techniques expressed with continuous variables, we need to adapt the above formalism to a discrete framework. In this regard, we consider the signal  $f$  to be a discrete sequence of length  $L$  such that  $f[n] = f(\frac{n}{L})$ , for  $n = 0, \dots, L-1$ , and that  $(g[n])_{n \in \mathbb{Z}}$  are the samples at  $\frac{n}{L}$  of the window, which is further truncated to be supported on  $\{-M, \dots, M\}$  such that  $2M+1 \leq N$ , where  $N$  is the number of frequency bins. In that context, the discrete STFT of  $f$  is defined by [15]

$$\begin{aligned} V_f^g[m, k] &:= \sum_{n \in \mathbb{Z}} f[n]g[n-m]e^{-2i\pi\frac{k(n-m)}{N}} \\ &= \sum_{n=-M}^{n=M} f[m+n]g[n]e^{-2i\pi\frac{kn}{N}} \\ &= \sum_{n=0}^{N-1} f[n+m-M]g[n-M]e^{-2i\pi\frac{k}{N}(n-M)}, \end{aligned} \quad (1.6)$$

with  $2M + 1 \leq N$ , where  $N$  is the number of frequency bins,  $m \in \{0, \dots, L - 1\}$ , and  $k \in \{0, \dots, N - 1\}$  corresponding to the frequency  $k \frac{L}{N}$ . The discrete transform is also invertible [15], if  $g[0] \neq 0$ ,

$$f[n] = \frac{1}{g[0]N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} V_f^g[n, k]. \quad (1.7)$$

Or if one assumes that  $f$  is  $L$ -periodic,  $g \neq 0$ ,

$$f[n] = \frac{1}{N\|g\|_2} \sum_{k=0}^{N-1} \sum_{m=n-M}^{n+M} V_f^g[m \bmod L, k] g[n-m] e^{i2\pi \frac{k(n-m)}{N}}. \quad (1.8)$$

Or, still assuming  $f$   $L$ -periodic, if  $\sum_{m=-M}^M g[m] \neq 0$ ,

$$f[n] = \frac{1}{N \sum_{m=-M}^M g[m]} \sum_{k=0}^{N-1} \sum_{m=n-M}^{n+M} V_f^g[m \bmod L, k] e^{i2\pi \frac{k(n-m)}{N}}. \quad (1.9)$$

In these last two cases, the STFT of  $f$  is also  $L$ -periodic with respect to time.

## 1.2. Reassignment and Synchrosqueezing

According to the uncertainty principle, for the STFT or the spectrogram, the energy is spread over the TF plane. To circumvent this effect, considering a signal  $f \in L^2(\mathbb{R})$  and a window  $g \in S(\mathbb{R})$ , one may define a TF point  $(\hat{t}_f(t, \eta), \hat{\omega}_f(t, \eta))$ , called centroid, on which to reallocate the energy of the STFT, provided  $V_f^g(t, \eta) \neq 0$ , and defined by [16]

$$\hat{t}_f(t, \eta) := t - \frac{1}{2\pi} \partial_\eta \arg V_f^g(t, \eta), \quad (1.10)$$

and

$$\hat{\omega}_f(t, \eta) := \frac{1}{2\pi} \partial_t \arg V_f^g(t, \eta). \quad (1.11)$$

In practice,  $\hat{t}_f(t, \eta)$  and  $\hat{\omega}_f(t, \eta)$  are computed through the definition of complex reassignment operators [17],

$$\tilde{t}_f(t, \eta) := t + \frac{V_f^{tg}(t, \eta)}{V_f^g(t, \eta)} \quad (1.12)$$

$$\tilde{\omega}_f(t, \eta) := \eta - \frac{1}{2i\pi} \frac{V_f^{g'}(t, \eta)}{V_f^g(t, \eta)}, \quad (1.13)$$

and then  $\hat{\omega}_f(t, \eta) = \Re\{\tilde{\omega}(t, \eta)\}$  and  $\hat{t}_f(t, \eta) = \Re\{\tilde{t}(t, \eta)\}$ , where  $\Re\{x\}$  denotes the real part of  $x$ . When  $f$  is a pure tone, namely  $f = Ae^{2i\pi\phi(t)}$  with  $\phi(t) = \omega t$ , then  $\hat{\omega}_f(t, \eta) = \omega$  for any  $(t, \eta)$  in the TF plane [17]. This means that  $\hat{\omega}_f$  gives the frequency  $\omega$  of the pure tone  $f$ . In the general case, where the amplitude  $A(t)$  and the frequency  $\phi'(t)$  depends on time,  $\hat{\omega}_f$  becomes an estimate of  $\phi'$ . It is also important to note that the operators  $\tilde{\omega}_f$  and  $\tilde{t}_f$  only depend on the computation of three STFTs and are therefore easy to compute. Then, assuming  $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  and  $\int_{\mathbb{R}} g \neq 0$ , the *reassignment method* (RM) associated with  $\hat{t}_f$  and  $\hat{\omega}_f$  is defined by [16],

$$RM(t, \eta) = \int_{\mathbb{R}^2} |V_f^g(\tau, \xi)|^2 \delta(t - \hat{t}_f(\tau, \xi)) \delta(\eta - \hat{\omega}_f(\tau, \xi)) d\tau d\xi \quad (1.14)$$

However, RM is not invertible because the phase is lost in the process, and therefore one may instead consider reallocating the STFT along the frequency axis only, which brings about the *Fourier-based synchrosqueezing transform* (FSST) whose principle is detailed hereafter.

### 1.2.1. Definition of the synchrosqueezing

FSST was introduced in [17, 18], and using both assumptions (1.5) and  $g \in S(\mathbb{R})$ , it is defined by,

$$T_f^g(t, \eta) := \int_{\mathbb{R}} V_f^g(t, \xi) \delta(\eta - \hat{\omega}_f(t, \xi)) d\xi. \quad (1.15)$$

Provided  $g(0) \neq 0$  one has, when  $f$  is analytic,

$$f(t) = \frac{1}{g(0)} \int_{\mathbb{R}^+} T_f^g(t, \eta) d\eta. \quad (1.16)$$

This transform leads to a sparser TFR than the STFT but is very sensitive to frequency modulation. In that context, to improve  $\hat{\omega}_f$ , the following complex modulation operator was introduced [17],

$$\tilde{q}_f(t, \eta) := \frac{\partial_t \tilde{\omega}_f(t, \eta)}{\partial_t \tilde{t}_f(t, \eta)} = \frac{1}{2i\pi} \frac{V_f^{g''}(t, \eta) V_f^g(t, \eta) - \left(V_f^{g'}(t, \eta)\right)^2}{V_f^{tg}(t, \eta) V_f^{g'}(t, \eta) - V_f^{tg'}(t, \eta) V_f^g(t, \eta)}. \quad (1.17)$$

Local modulation is given by  $\hat{q}_f(t, \eta) := \Re\{\tilde{q}_f\}$ . An illustration of  $\hat{q}_f$  is given in Figure

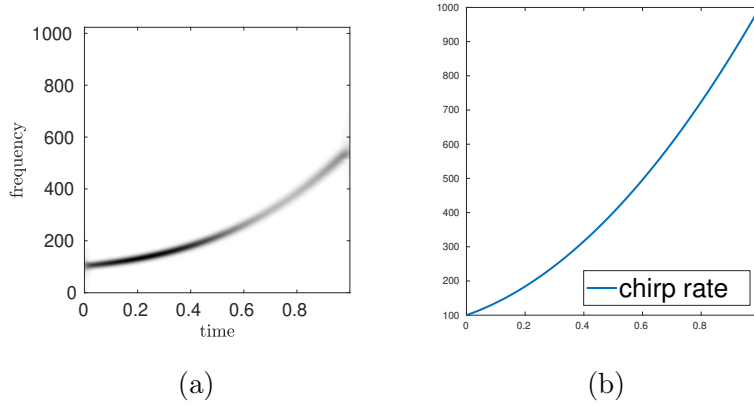


Figure 1.1.: (a) STFT of  $f(t) = \exp(2i\pi(50(t^4 + t^3 + t^2) + 100t))$ ; (b) instantaneous chirp rate  $CR(t) = 600t^2 + 300t + 100$

1.1. Furthermore, for a linear chirp with Gaussian modulated amplitude  $f$ , namely  $f(t) = A(t)e^{2i\pi\phi(t)}$  where  $\log(A(t))$  and  $\phi(t)$  are second order polynomials,  $\hat{q}_f(t, \eta)$  is the *chirp rate* (CR), namely  $\hat{q}_f = \phi''(t)$  [17, 19]. Using  $\tilde{q}_f$ , the second order frequency reallocation operator was introduced,

$$\hat{\omega}_f^{[2]}(t, \eta) = \Re \{ \tilde{\omega}_f(t, \eta) - \tilde{q}_f(t, \eta)(t - \tilde{t}_f(t, \eta)) \}, \quad (1.18)$$

which verifies  $\hat{\omega}_f^{[2]}(t, \eta) = \phi'(t)$ . The *second order Fourier-based synchrosqueezing transform* (FSST2) can then be defined [17, 19],

$$T_f^{g,2}(t, \eta) := \int_{\mathbb{R}} V_f^g(t, \xi) \delta(\eta - \hat{\omega}_f^{[2]}(t, \xi)) d\xi. \quad (1.19)$$

Note that to evaluate  $\tilde{q}_f$  and  $\hat{\omega}_f^{[2]}$  only requires to compute five STFTs. The *Fourier based synchrosqueezing* has been extended to the  $N$ -th order [20], denoted  $T_f^{g,N}$ , which we will detail and use only in Chapter 4. Note that the *synchrosqueezing transform* (SST) was initially introduced in the wavelet framework [21], but in the context of this thesis, we will stick to the STFT formalism.

### 1.3. Multicomponent signals

Many real world non-stationary signals can be described using the formalism of *multicomponent signals* (MCSs). For instance, real world audio signals such as piano, clarinet [22], bird

songs [23], speech [24] but also signals from electrocardiograms [25–27], thoracic and abdominal movements [28], and underwater propagation [29].

A MCS  $f$  containing  $P$  modes is defined by,

$$f(t) = \sum_{p=1}^P f_p(t), \quad (1.20)$$

with  $f_p(t) = A_p(t)e^{2i\pi\phi_p(t)}$  where  $A_p$  is the *instantaneous amplitude* and  $\phi_p$  is the *instantaneous phase*. Then, the *instantaneous frequency* (IF) of  $f_p$  is given by  $\phi'_p$ . When the IF of the modes are well separated in frequency, curves reflecting the energy of each mode can be observed in the TFR, as illustrated in Figure 1.2.

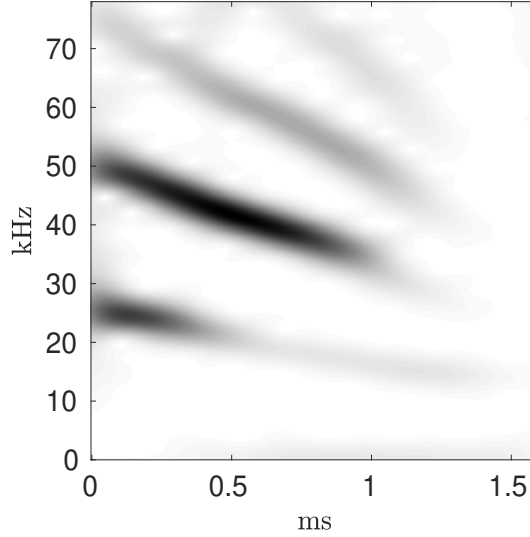


Figure 1.2.: Spectrogram of a bat echolocation signal.

Therefore, MCS are commonly analyzed in time and frequency [18].

Then, its ideal TFR is given by,

$$ITF_f(t, \eta) = \sum_{p=1}^P A_p(t)\delta(\eta - \phi'_p(t)), \quad (1.21)$$

where  $\delta$  is the Dirac distribution. It is said ideal in the sense that it is perfectly localized, both in time and frequency. The SST aims towards this representation and is close to it when the signal meets the assumptions the transform is designed for. An illustration of *ideal time frequency representation* (ITF) is given in Figure 1.3.

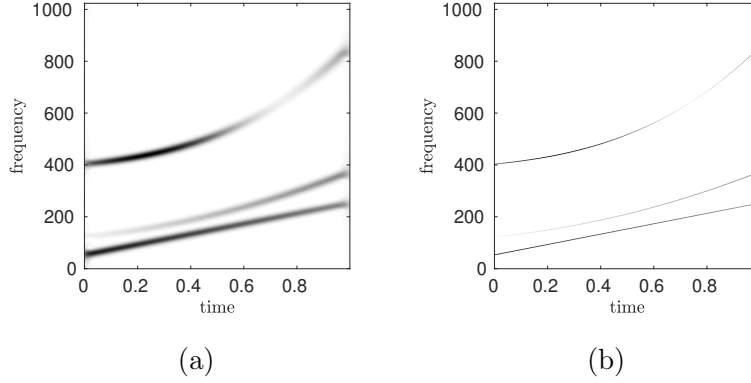


Figure 1.3.: (a): STFT of a MCS, (b) ITF of the signal of (a).

### 1.3.1. Mode retrieval

When considering a MCS  $f$ , there is no general formula to recover each of its modes. The process of computing an approximation of a mode is called *mode retrieval* (MR). Existing techniques use various TFRs such as the *continuous wavelet transform*, as for instance in [30], or in the context of synchrosqueezing [21]. When STFT is used, one can take advantage of its linearity, namely,

$$V_f^g(t, \eta) = \sum_{p=1}^P V_{f_p}^g(t, \eta), \quad (1.22)$$

as for instance, in [15] to perform MR based on hard-thresholding.

Then, the retrieval of  $f_p$  is performed using one of the approximation formulae of  $V_{f_p}^g(t, \eta)$  (1.3), (1.4) or (1.5), depending on what assumptions are made on  $f$  and  $g$ .

FSST was also used for that matter, for instantaneous frequency estimation [31], for non-stationnary modes [32], a thorough mathematical study being available in [33].

When the separation in the TF plane is difficult, adaptive approaches can also be considered, namely the analysis window may vary in time or frequency, as in the case of adaptive wavelets based SST [34], or in the case of adaptive STFT based SST [35] which has recently been used, in [36] for multicomponent signal separation.

For the sake of completion, we note that other techniques, not based on TFR, are also used, as those based on chirplet [37] or ridglet [38] transforms.



## 1.4. Outline of the Following Chapters

In Chapter 2, we will carry out TF analysis of a MCS affected by noise in order to perform MR. In our approach, we are first going to analyze the TF plane to associate TF regions to each mode. This task is called *ridge detection* (RD) as those regions often look like ridges due to the amount of energy they contain [39]. Then, we perform a local analysis in the vicinity of the ridges in order to approximate the corresponding modes using a linear chirp based model. This enables us to express the STFT of the mode in a closed form and then to define a numerical procedure to perform MR.

In Chapter 3, we will perform a TF analysis of the interference of MCSs. We will first characterize an interference pattern called *time frequency bubble* (TFB) theoretically and subsequently introduce a technique for their identification in the TF plane, which we validate numerically. At end of the chapter, we will present perspective using a different TF approach to identify interference based on the spectrogram zeros. For this purpose, we study the energy distribution related to the Voronoi cell, which will be continued in a near future.

In Chapter 4, we will recall the definition of *high order Fourier-based synchrosqueezing transform* (FSSTN), and then, we will show how to make an adaptive choice of its order based on an energy based criterion, and validate this technique numerically.

In Chapter 5, considering that FSST2 is relevant in very noisy situations, we will study its associated CR estimator. We illustrate that the errors introduced by the noise have an oscillatory nature and propose a technique to filter out most of them. We will finally validate its accuracy on numerical experiments and its stability on real data.

In Chapter 6, we will study a TF approach for *heart rate* (HR) estimation on *electrocardiogram* (ECG) signals. Assuming that an ECG signal can be represented by a harmonic model, we will propose a TF technique for HR estimation, and show that the choice of the TFR has a huge impact on the accuracy. Finally, we will introduce an algorithm based on FSST for the estimation of the HR and evaluate it on synthetic and real ECG signals.

## 1.5. List of published articles linked to the contributions of this thesis

In the context of Chapter 2,

- 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
- N. Laurent and S. Meignen, “A novel time-frequency technique for mode retrieval based on linear chirp approximation”, *IEEE Signal Processing Letters*, vol. 27, pp. 935–939, 2020

In the context of Chapter 4,

- N. Laurent and S. Meignen, “A new adaptive technique for multicomponent signals reassignment based on synchrosqueezing transform”, in *2022 30th European Signal Processing Conference (EUSIPCO)*, IEEE, 2022, to appear

In the context of Chapter 5,

- N. Laurent *et al.*, “On local chirp rate estimation in noisy multicomponent signals: With an application to mode reconstruction”, *IEEE Transactions on Signal Processing*, vol. 70, pp. 3429–3440, 2022

In the context of Chapter 6,

- N. Laurent *et al.*, “A novel algorithm for heart rate estimation based on synchrosqueezing transform”, in *2021 29th European Signal Processing Conference (EUSIPCO)*, IEEE, 2021, pp. 1286–1290



## 2. Mode Retrieval Based on Robust Ridge Detection and Linear Chirp Approximation

### 2.1. Introduction

In this chapter, we perform TF analysis to design a MR technique robust to noise. For this task, we are going to rely on curves called *ridges* which approximate the IF of the modes of a MCS [45, 46]. The detection of those curves is preformed by a RD.

In a first part of this chapter, we consider the ridge detection in the case of noise. In the literature, there exist RD based on different TFR, for instance the continuous wavelet transform [47], the Wigner Ville transform [48], or on STFT [49]. In our approach, we use TF regions with high energy as an indicator of signal information [50, 51]. The technique we detail is designed to be robust to noise, and therefore, it does not assume that the ridge is continuous. Instead of trying to follow a ridge, it uses structures based on *local maxima along the frequency axis* (LMMFs).

In the second part of this chapter, we design a MR technique based on a RD for noisy signals. We show that even with perfect ridge detection, the quality of the reconstruction cannot be guaranteed. One possible approach is to perform hard thresholding and proceed to the reconstruction of the signal [15]. However, TF regions related to the signal are still affected by noise. Our approach tries to limit the effect of the noise in the signal domain by using a linear chirp model.

At the end of this chapter, we finally perform numerical validation of the techniques and draw conclusions on their behaviors.

---

This chapter is based on the work published in [41] and [40].

As we will always consider noisy situations, let us introduce the noisy signal,

$$\tilde{f}(t) := f(t) + \varepsilon(t), \quad (2.1)$$

where  $f$  is the MCS and  $\varepsilon$  is a complex, zero mean, white Gaussian noise with variance  $\sigma_\varepsilon^2$ . We assume  $\varepsilon$  to be independently and identically distributed (i.i.d) with respect to time and that its real and imaginary parts are independent as well. In the remainder of this chapter, we will also use a discrete setting for  $\tilde{f}$ , by considering it as a sequence of length  $L$  with  $\tilde{f}[n] = \tilde{f}(\frac{n}{L})$ .

## 2.2. Basics of Ridge Detection

Performing ridge detection on  $\tilde{f}$  consists of associating TF curves to the modes of  $\tilde{f}$ . The objective of this section is to understand basics of what makes a RD robust to noise. Therefore, we first study the effect of  $\varepsilon$  on non-robust RDs, the focus being put on a very commonly used one, which we call *simple ridge detector* (S-RD), in the sequel and that we introduce hereafter.

### 2.2.1. Simple STFT-Based Ridge Detection

The RD introduced by [47] is based on finding the location of the maxima of energy in the spectrogram. It consists of determining the  $P$  ridges associated with the modes in the TF plane by computing:

$$\max_{\mathbf{\Gamma}} \sum_{p=1}^P \sum_{n=0}^{L-1} |V_{\tilde{f}}^g[n, \Gamma_p[n]]|^2 - \alpha(\Delta^1 \Gamma_p[n])^2 - \beta(\Delta^2 \Gamma_p[n])^2, \quad (2.2)$$

with  $\mathbf{\Gamma} = (\Gamma_p)_{p=1, \dots, P}$ ,  $\alpha$  and  $\beta$  both positive, and in which  $\Delta^1 \Gamma_p[n]$  (resp.  $\Delta^2 \Gamma_p[n]$ ) is the first order (resp. second) order finite differences approximating  $\phi_p''[n]$  (resp.  $\phi_p^{(3)}[n]$ ). To consider penalization terms is however not relevant at low noise level [52], and these are therefore often left apart. Alternatively, one can use a bound  $B_f$  on the frequency modulation of the modes to extract a first ridge, and then replace (2.2) by a *peeling algorithm*. In a nutshell, a first ridge is extracted as follows [49]:

$$\max_{\Gamma_1} \sum_{n=0}^{L-1} |V_{\tilde{f}}^g[n, \Gamma_1[n]]|^2, \quad \text{s.t. } |\Delta^1 \Gamma_1[n]| \leq B_f. \quad (2.3)$$

In practice, to compute  $\Gamma_1$ , one initializes  $\Gamma_1[n_0]$  to  $k_0 := \underset{0 \leq k \leq N-1}{\operatorname{argmax}} |V_{\tilde{f}}^g[n_0, k]|$ . Then, to define  $\Gamma_1$  on  $\{n_0 + 1, \dots, L - 1\}$ , one uses the following induction principle starting from  $n = n_0$ :

$$\Gamma_1[n + 1] := \underset{k}{\operatorname{argmax}} \left\{ |V_{\tilde{f}}^g[n + 1, k]|, \text{ s.t. } |k - \Gamma_1[n]| \leq B_f \right\}. \quad (2.4)$$

The same principle is applied on  $\{0, \dots, n_0 - 1\}$ , starting from  $n = n_0$  and replacing  $n + 1$  by  $n - 1$  in (2.4). Finally, the procedure is run again starting from other initial time indices to define other candidates for  $\Gamma_1$ , and the ridge finally kept among all the candidates is the one maximizing the energy in the TF plane, i.e.  $\sum_n |V_{\tilde{f}}^g[n, \Gamma_1[n]]|^2$ .

Then, to compute the next ridges, one introduces the following set, for  $p > 1$ :

$$\mathcal{S}_p(n) = \{0 \leq k \leq N - 1, \text{ s.t. } |k - \Gamma_q[n]| \geq \Delta, \ 1 \leq q < p\}, \quad (2.5)$$

and then the  $p^{\text{th}}$  ridge is computed using the same principle as for  $\Gamma_1$  except the initialization index is defined as  $k_0 := \underset{k \in \mathcal{S}_p(n)}{\operatorname{argmax}} |V_{\tilde{f}}^g[n_0, k]|$  and then (2.4) is replaced by:

$$\Gamma_p[n + 1] := \underset{k \in \mathcal{S}_p(n+1)}{\operatorname{argmax}} \left\{ |V_{\tilde{f}}^g[n + 1, k]|, \text{ s.t. } |k - \Gamma_p[n]| \leq B_f \right\}. \quad (2.6)$$

Such a technique is iterated until  $P$  ridges  $(\Gamma_p)_{p=1, \dots, P}$  are extracted. There are however two strong limitations to S-RD:

- \* Each ridge is built assuming that the selection of the maximum around the current position provides with a continuous curve on which the coefficients are associated with signal information. This assumption does not hold when the noise is too strong. Indeed, in such a case, zeros of STFT may appear at TF locations corresponding the IFs of the modes. This is illustrated on a linear chirp in Figure 2.1 in which we display the magnitude of the LMMFs associated with the three largest STFT modulus maxima at each time instant, along with the true IF (*signal to noise ratio* (SNR) = -10 dB).
- \* The jumps allowed between two successive time indices depend on the modulation parameter  $B_f$  which is fixed a priori, and thus this method does not adapt to the local variations of the frequency modulation of the modes.

We next recall the RD proposed in [49], where the chaining of the coefficient is made more adaptive by removing the modulation parameter  $B_f$  and exploiting the modulation

operator  $\hat{q}_{\hat{f}}$  instead.

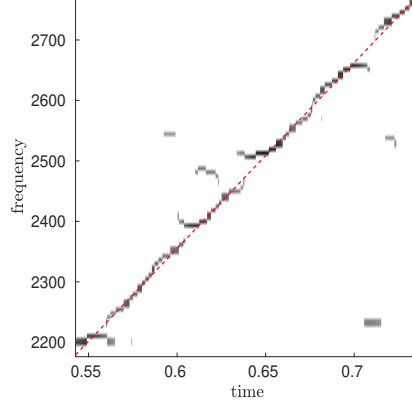


Figure 2.1.: LMMFs corresponding to the three largest STFT modulus maxima for each time instant along with the true IF (dashed line) of the linear chirp (SNR= -10 dB )

### 2.2.2. Modulation Based Ridge Detection

To circumvent the lack of adaptivity of S-RD to the local variations of the frequency modulation of the modes, a novel approach called *modulation based ridge detector* (MB-RD) was proposed in [49]. In a nutshell, this approach extracts the first ridge, MB-RD uses the same induction principle as S-RD, but considers the complex modulation operator  $\hat{q}_{\hat{f}}$  defined in Section 1.2 instead of  $B_f$ , namely,

$$\Gamma_1[n+1] := \underset{k}{\operatorname{argmax}} \left\{ |V_{\hat{f}}^g[n+1, k]|, |k - \Gamma_1[n] - \hat{q}_{\hat{f}}[n, \Gamma_1[n]] \frac{N}{L^2}| \leq C \right\}, \quad (2.7)$$

the user-defined constant  $C$  compensating for potential local frequency modulation estimation errors. MB-RD is proved to be slightly sensitive to  $C$  when the noise level is low [49] (the simulations in that paper only considered a SNR larger than 0 dB), but the estimator  $\hat{q}_{\hat{f}}$  is not robust enough when the noise is too strong, which leads to bad results.

Furthermore, both techniques are based on the assumption that the IF of a mode at each time index can be associated with a continuous sequence of energetic TF coefficients, which may not be the case in heavy noise situations. Another limitation of S-RD and MB-RD is that they build the ridges one after the other using the peeling algorithm recalled in Section 2.2.1: if the ridge detection fails for one mode, it will also fail for the

next ones. To deal with all these issues, the concept of *relevant ridge portions* (RRPs) is introduced in the following section, and subsequently used to define a new RD, for which the approach significantly differs from S-RD and MB-RD.

## 2.3. Definition of a New Robust Ridge Detector

The objective of this section is to define a RD robust to noise. Its algorithm can be divided in three major steps. The first step consists of identifying energetic chains of coefficients belonging to the signal, called RRP. The second step associates RRP that have connected basins of attraction. Such a basin is a TF regions, where all the coefficients points towards the same RRP. The last step consists of two nested iterative procedures. The first one selects a combination of RRP to associate with each mode. Each of them represents initial TF coefficients that can be trusted to be associated with one mode of the signal. The second one, uses the initialization given by the first procedure to associate a TF curve with each mode using an iterative procedure based on weighted spline. We name the solution given by this algorithm *relevant ridge portion based ridge detector* (RRP-RD).

### 2.3.1. Definition of Relevant Ridge Portions

To have a robust RD, we first define a structure that persists even when noise strongly impacts the signal. For that purpose, we use  $\hat{q}_{\tilde{f}}$  to define a stable direction in the TF plane between two neighboring LMMFs. Denoting a generic LMMF by  $[n, m[n]]$ , we introduce,

**Definition 2.3.1.** *Let  $[n, m[n]]$  and  $[n+1, m[n+1]]$  be two LMMFs, they satisfy relation  $([n, m[n]] \sim [n+1, m[n+1]])$  if and only if,*

$$\begin{aligned} m[n+1] &= \underset{k}{\operatorname{argmin}} \left\{ \left| k - m[n] - \hat{q}_{\tilde{f}}[n, m[n]] \frac{N}{L^2} \right|, \text{ s.t. } [n+1, k] \text{ LMMF} \right\} \\ m[n] &= \underset{k}{\operatorname{argmin}} \left\{ \left| m[n+1] - k - \hat{q}_{\tilde{f}}[n+1, m[n+1]] \frac{N}{L^2} \right|, \text{ s.t. } [n, k] \text{ LMMF} \right\}. \end{aligned} \quad (2.8)$$

If  $[n, m[n]] \sim [n+1, m[n+1]]$ ,  $\hat{q}_{\tilde{f}}$  is a relevant operator giving the direction from  $[n, m[n]]$  to  $[n+1, m[n+1]]$  and vice-versa, meaning that  $\hat{q}_{\tilde{f}}$  gives a stable orientation. A *ridge portion* (RP) is then defined by extending relation  $\sim$  to any time instants  $n, n_0$



that are not necessarily neighbors. Assuming, without loss of generality, that  $n > n_0$ , we define relation  $\leftrightarrow$ ,

**Definition 2.3.2.** Let  $[n, m[n]]$  and  $[n_0, m[n_0]]$  be two LMMFs such that  $n \geq n_0 + 1$ , then define:

$$([n_0, m[n_0]] \leftrightarrow [n, m[n]]) \Leftrightarrow \begin{cases} \exists [n_0 + 1, m[n_0 + 1]], \dots, [n - 1, m[n - 1]] \text{ LMMFs} \\ \forall i = n_0, \dots, n - 1, [i, m[i]] \sim [i + 1, m[i + 1]] \end{cases} \quad (2.9)$$

A RP  $\mathcal{R}$  containing LMMF  $[n_0, m[n_0]]$ , is finally defined by:

$$\mathcal{R}[n_0, m[n_0]] = \{[n, m[n]], \text{s.t. } [n, m[n]] \leftrightarrow [n_0, m[n_0]]\}. \quad (2.10)$$

Note that, if  $[n, m[n]] \in \mathcal{R}[n_0, m[n_0]]$  then  $\mathcal{R}[n, m[n]] = \mathcal{R}[n_0, m[n_0]]$ .

The definition of RPs allows considering the discontinuities in the ridges associated with a mode. However, to separate the noise from the signal, a criterion needs to be defined. Using the assumptions made on the noise  $\varepsilon$  [53], we have that,

$$\text{Var}(\Re\{V_\varepsilon^g[n, k]\}) = \text{Var}(\Im\{V_\varepsilon^g[n, k]\}) = \sigma_\varepsilon^2 \|g\|_2^2.$$

Then, remarking that  $\frac{|V_\varepsilon^g|^2}{\sigma_\varepsilon^2 \|g\|_2^2}$  is  $\chi_2$  distributed with two degrees of freedom and assuming the variance of the noise  $\sigma_\varepsilon^2$  is known, the probability that  $|V_\varepsilon^g[n, k]| \geq \beta \sigma_\varepsilon \|g\|_2$  is smaller than 10% if  $\beta = 2$ , and 1% if  $\beta = 3$ . So, by considering only the LMMFs  $[n, m[n]]$  such that  $|V_{\hat{f}}^g[n, m[n]]| \geq \beta \sigma_\varepsilon \|g\|_2$  with  $\beta \geq 2$  one removes many LMMFs corresponding to noise. In practice,  $\sigma_\varepsilon$  is unknown, but can be approximated using the estimate  $\gamma = \sigma_\varepsilon \|g\|_2$  defined by [54, 55],

$$\hat{\gamma} = \frac{\text{median} \left| \Re \left\{ V_{\hat{f}}^g[n, k] \right\}_{n,k} \right|}{0.6745},$$

in which median represents the median of the coefficients. One can then set a meaningful threshold on the STFT, taking into account the above probabilities, and considering

$$\mathcal{S}(\beta) = \left\{ [n, k], |V_{\hat{f}}^g[n, k]| \geq \beta \hat{\gamma} \right\}. \quad (2.11)$$

Using  $\mathcal{S}(\beta)$ , we are able to define a RRP, that is a RP satisfying relation  $\leftrightarrow$  defined by,

**Definition 2.3.3.** Let  $[n, m[n]]$  and  $[n_0, m[n_0]]$  be two LMMFs such that  $n \geq n_0 + 1$ , then define:

$$[n_0, m[n_0]] \rightsquigarrow [n, m[n]] \Leftrightarrow \begin{cases} \exists [n_0 + 1, m[n_0 + 1]], \dots, [n - 1, m[n - 1]] \text{ LMMFs} \\ \forall i = n_0, \dots, n - 1, \begin{cases} [i, m[i]] \sim [i + 1, m[i + 1]] \\ [i, m[i]] \in \mathcal{S}(\beta) \end{cases} \end{cases} \quad (2.12)$$

Any LMMF belonging to a RRP is called a relevant LMMF. To properly separate signal from noise, a study of the parameter  $\beta$  is required.

### Fixing the parameter $\beta$ for the RRP

A crucial issue is then how to fix the parameter  $\beta$ . For that purpose, we numerically investigate the existence of relevant LMMFs around the true IF when  $\beta$  varies. To define a meaningful interval around the true IF, we first compute analytical expressions of the STFT assuming a local linear chirp approximation for the modes. To this end, we set  $g$  to be a Gaussian window for two reasons:

- \* First because it simplifies the computation of  $\hat{q}_{\tilde{f}}$ .
- \* Second because the expression of the STFT of a signal that can be locally approximated by a linear chirp is particularly simple [19].

Indeed, let  $g$  be the discrete Gaussian window  $g[n] = e^{-\pi \frac{n^2}{\sigma^2 L^2}}$ , to increase the readability of the TFR, we choose the value of  $\sigma$  whose associated STFT modulus minimizes the Rényi entropy and should therefore provide good localization. Such a choice for  $\sigma$  is proved to minimize interference between the modes in the TF plane [56, 57]. With such a window, if  $f$  can be locally approximated by a linear chirp with constant amplitude  $A$ , i.e.  $f[n] \approx Ae^{2i\pi\phi[n]}$ , with  $\phi$  a second order polynomial, one has [19],

$$|V_f^g[n, k]| \approx AL\sigma(1 + \sigma^4\phi''[n]^2)^{-\frac{1}{4}} e^{-\pi \frac{\sigma^2(k\frac{L}{N} - \phi'[n])^2}{1 + \sigma^4\phi''[n]^2}}, \quad (2.13)$$

which is a Gaussian function with standard deviation,

$$\Delta_{LC}[n] = \frac{1}{\sqrt{2\pi}\sigma} \sqrt{1 + \sigma^4\phi''[n]^2}. \quad (2.14)$$

For  $V_f^g$ , the LMMF associated with the largest STFT modulus maximum at time index  $n$  has its ordinate in the interval  $\mathcal{I}[n] = [(\phi'[n] - \Delta_{LC}[n])\frac{N}{L}, (\phi'[n] + \Delta_{LC}[n])\frac{N}{L}]$ . When some noise is added, this is no longer true. In the following experiment, we assume  $\phi'$

and  $\phi''$  are known, and assess the proportion corresponding to the number of relevant LMMFs with ordinate in  $\mathcal{I}[n]$  when  $n$  varies, namely:

$$\mathcal{P}(\beta) = \frac{\#\{[n, m[n]] \in \mathcal{S}(\beta) \text{ with } m[n] \in \mathcal{I}[n]\}}{L}, \quad (2.15)$$

where  $\#X$  denotes the cardinal of the set  $X$ . Using numerical validation, we show how to set  $\beta$  as high as possible to filter out the noise while keeping  $\mathcal{P}(\beta)$  as high as possible. Such a value for  $\beta$  enables to keep a maximum of relevant LMMFs in  $\mathcal{I}$  which will help us associate them to the same modes by connecting their corresponding RRP.

We carry out the validation of the choice of  $\beta$  for the three signals of Figure 2.2 (a), which, from top to bottom, are more and more modulated, and get the results of Figure 2.2 (b). These latter correspond to experiments with an input SNR of -10 dB, meaning that the noise is very strong. The quantity  $\mathcal{P}(\beta)$  is almost constant for  $\beta$  between 0 and 2, before decreasing rapidly. We are therefore going to choose  $\beta = 2$  assuming that there will be enough RRP to represent the ridge of each mode. From now on,  $\beta$  equals 2 unless mentioned otherwise.

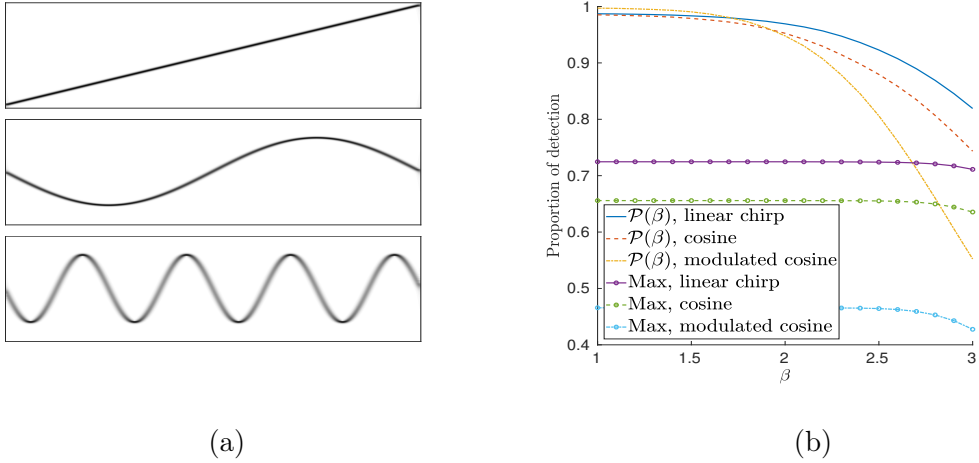


Figure 2.2.: (a): from top to bottom: STFT of a linear chirp, of a signal with cosine phase, and of a signal with oscillatory phase; (b):  $\mathcal{P}(\beta)$  for these three signals when the input SNR equals -10 dB as well as the proportion of time indices  $n$ , at which the relevant LMMF with ordinate in  $\mathcal{I}[n]$  corresponds to the largest STFT modulus maximum (curves denoted by "Max, linear chirp", "Max, cosine", and "Max, modulated cosine"). The results are averaged over 100 noise realizations.

### 2.3.2. Grouping RRP's Based on Basins of Attraction

In this subsection, the objective is to group RRP's using the connection of their *basin of attractions* (BAs) to finally assign these groups to one mode of the signal. For that purpose, we will use the reallocation vectors  $(\hat{t}_{\hat{f}}[n, k], \hat{\omega}_{\hat{f}}[n, k])$  defined in 1.2. Indeed, a BA associated with the RRP  $\mathcal{R}_i$  can be defined by,

$$\mathcal{B}_i := \left\{ [n, k]; \quad \underset{[x, y] \in \text{RRPs}}{\operatorname{argmin}} \quad \|(\hat{t}_{\hat{f}}[n, k], \hat{\omega}_{\hat{f}}[n, k]) - [x, y]\| \in \mathcal{R}_i \right\}. \quad (2.16)$$

However, many BAs are associated with noise, and therefore we impose more conditions to keep the most energetic part of  $\mathcal{B}_i$ . To do so, we consider a *hard thresholding* (HT) approach,

$$\mathcal{B}_i^{HT} = \begin{cases} \mathcal{B}_i \cap \mathcal{S}(2) & \text{if } \mathcal{R}_i \cap \mathcal{S}(3) \neq \emptyset \\ \emptyset & \text{otherwise,} \end{cases} \quad (2.17)$$

In short,  $\mathcal{B}_i^{HT}$  is more likely related to the signal than  $\mathcal{B}_i$ , since it contains at least one coefficient in  $\mathcal{S}(3)$ .

We then gather together connected  $\mathcal{B}_i^{HT}$ s in the TF plane to obtain a set of larger TF regions which are denoted by  $\{(\mathcal{C}_j^{HT})_j\}$  in the sequel.

### 2.3.3. New RD Definition

In this subsection, we detail how we use the gathered BAs to associate a TF a curve to each mode of the signal. This is done using Algorithm 1 which we detail here. First, in order to avoid mixing ridges associated with different modes, the algorithm looks for  $P$  elements in  $\{(\mathcal{C}_j^{HT})_j\}$  coexisting in time. Indeed, ridges that coexist in time cannot belong to the same mode. In this approach, we consider the  $P$ -tuple coexisting on the longest set of time indices first. Therefore, Algorithm 1 starts with the ranking procedure “**sort-by-coexistence**”, to rank  $P$ -tuples of  $\{(\mathcal{C}_j^{HT})_j\}$  from the ones coexisting the longest to the shortest, and we denote this ranking by  $(\{\mathcal{C}_{p=1, \dots, P}^\kappa\})_\kappa$ . The first tuple of this ranking, namely  $(\mathcal{C}_p^0)_{p=1, \dots, P}$ , should contain one long ridge for each mode  $p$ . Because  $\{(\mathcal{C}_j^{HT})_j\}$  is finite,  $(\{\mathcal{C}_{p=1, \dots, P}^\kappa\})_\kappa$  is a finite sequence, and we denote the last element index by  $\kappa_{max}$ .

Next, we iterate over this ranked set to associate possible solutions for the ridge detection. At the  $\kappa$ -th iteration, where  $\kappa \in \{0, \dots, \kappa_{max}\}$ , we consider that the coefficients defined by  $\mathcal{A}_p^\kappa = (\mathcal{C}_p^0 \cup \mathcal{C}_p^1 \cup \dots \cup \mathcal{C}_p^\kappa) \cap \mathcal{S}(3)$  represent robust signal information that should be used as a basis to define the ridge of the mode  $p$ . In Algorithm 1, the procedure “**fit-spline**” defines an initial ridge using the following weighted spline approximation

formula,

$$s_{p,\text{initial}}^\kappa = \underset{s}{\operatorname{argmin}} \left[ (1 - \lambda) \sum_{[n,m[n]] \in \mathcal{A}_p^\kappa} \left( m[n] \frac{L}{N} - s\left(\frac{n}{L}\right) \right)^2 |V_{\tilde{f}}^g[n, m[n]]| + \lambda \int_0^1 (s''(t))^2 dt, \right]. \quad (2.18)$$

Then, “**fit-spline**” further iterates, by taking into account the elements in  $\{(C_j^{HT})\}$  that  $s_{p,\text{initial}}^\kappa$  crosses in (2.18). It continues to iterate until the spline stabilizes, meaning no new element are crossed in  $\{(C_j^{HT})\}$ . We denote the stabilized spline by  $s_p^\kappa$ .

---

**Algorithm 1:** New RD

---

**Input** :  $\{C_j^{HT}\}, P, S(3)$   
 $(C_p^\kappa) = \text{sort-by-coexistence}(\{C_j^{HT}\})$   
**for**  $\kappa = 0, \dots, \kappa_{\max}$  **do**  
    **for**  $p = 1, \dots, P$  **do**  
         $\mathcal{A}_p^\kappa = (C_p^0 \cup \dots \cup C_p^\kappa) \cap S(3)$   
         $s_p^\kappa = \text{fit-spline}(\mathcal{A}_p^\kappa, \{C_j^{HT}\})$   
        Compute energy  $E_p^\kappa$  using (2.20)  
    **end**  
**end**  
Compute index  $\kappa^{fin}$  using (2.19)  
Define the final ridges as  $s_p^{fin} := s_p^{\kappa^{fin}}$

---

Finally, we define the index  $\kappa^{fin}$  of the solution by

$$\kappa^{fin} := \underset{\substack{\kappa \text{ s.t. } (s_p^\kappa)_p \\ \text{are not crossing.}}}{\operatorname{argmax}} \sum_{p=1}^P E_p^\kappa, \quad (2.19)$$

where  $E_p^\kappa$  corresponds to the energy associated with mode  $p$  and is defined by,

$$E_p^k := \sum_{\substack{[n,m[n]] \in \mathcal{A}_p^k, \\ [n, \lfloor s_p^k[n] \frac{N}{L} \rfloor] \in \mathcal{C}_p^k}} |V_{\tilde{f}}^g[n, m[n]]|, \quad (2.20)$$

in which  $\lfloor X \rfloor$  denotes the nearest integer to  $X$ . In short,  $\{s_p^{fin}\}$  is chosen among the  $\kappa_{\max} + 1$  solutions. It maximizes an energy and contain no crossing splines. We denote this ridge detector RRP-RD.

We display on the first row of Figure 2.3 the modulus of the STFT of noisy two-mode signals made either of two linear chirps, of two modes with cosine phase, or of a

linear chirp plus an exponential chirp. In each case, we consider complex white Gaussian noise and the input SNR equals  $-10$  dB. On the second row of Figure 2.3, we display  $(s_p^{fin})_{p=1,2}$  computed with the optimization procedure just described. We notice that RRP-RD seems to be well adapted to deal with MCSs in the presence of heavy noise regardless of the modulation of the modes.

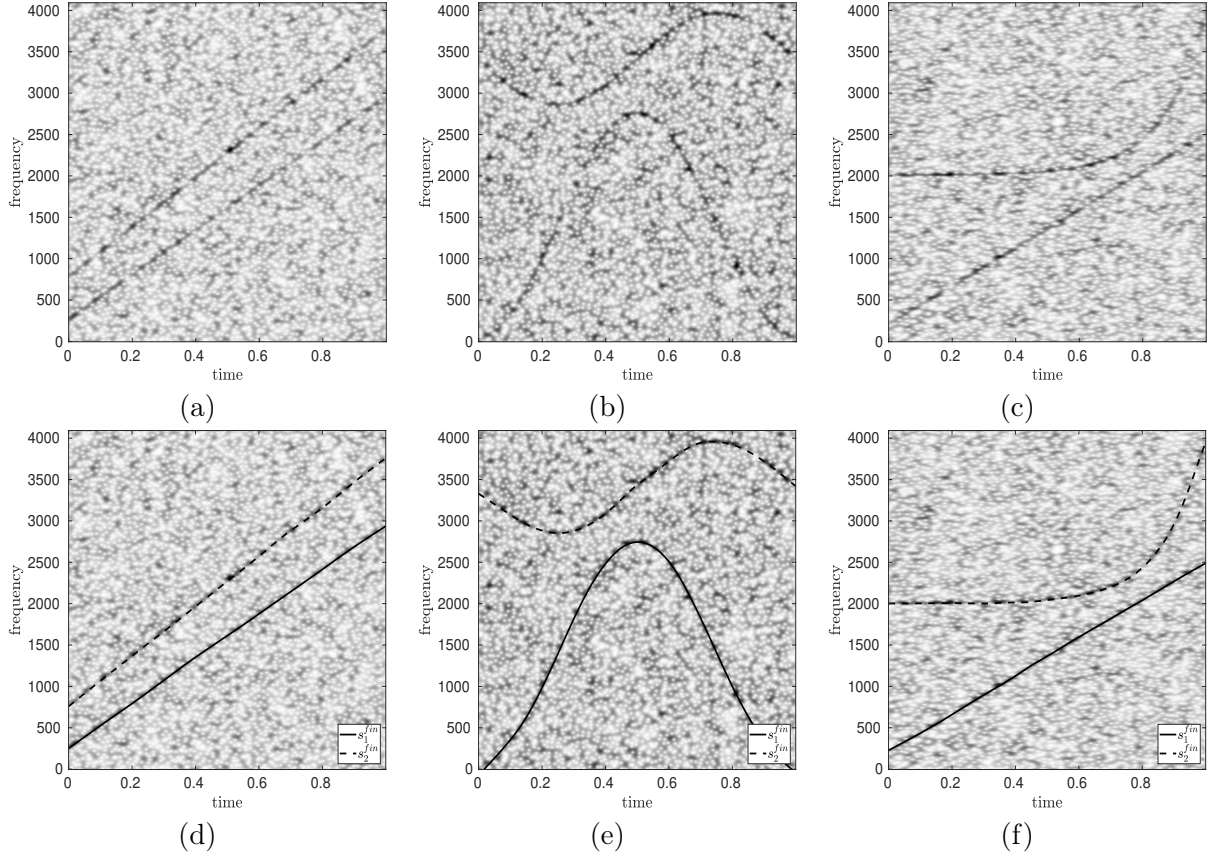


Figure 2.3.: (a): STFT modulus of two noisy linear chirps ( $\text{SNR} = -10$  dB,  $\sigma = 0.0188$ ); (b): STFT modulus of two noisy modes with cosine phase, with different modulation ( $\text{SNR} = -10$  dB,  $\sigma = 0.0175$ ); (c): STFT modulus of a signal made of a linear chirp and a mode with exponential phase ( $\text{SNR} = -10$  dB,  $\sigma = 0.0241$ ); (d):  $(s_p^{fin})_{p=1,2}$  computed for the signal displayed in (a); (e): same as (d) but for the signal whose STFT modulus is displayed in (b); (f): same as (d) but for the signal whose STFT modulus is displayed in (c).

#### 2.3.4. Comparison of RRP-RD, S-RD and MB-RD on Simulated Signals

In this subsection, we draw conclusions on the efficiency of RRP-RD compared with S-RD and MB-RD. In Figure 2.4 we show numerical results on the comparison of these

algorithms when the input SNR varies between -10 and 0 dB for the three signals of Figure 2.3. In particular, we here compare how close the ridge provided by any of the methods is close to the true IF for all the modes. For that purpose, we define the SNR by,

$$SNR(x, y) = 20 \log_{10}(\|x\|/\|y - x\|), \quad (2.21)$$

where  $y$  is an approximation of  $x$ . The first thing to notice is that RRP-RD always gives a significantly better SNR, from which we deduce that RRP-RD is indeed more robust to noise. Furthermore, the results are better when the smoothing parameter  $\lambda$  is larger. This can be explained by the fact that for those three signals, the IF is smooth, and a large  $\lambda$  enables to reduce the impact of the irregularities brought by strong noise.

We notice that the results for the two linear chirps of Figure 2.3 (a) are the best in terms of SNR, as displayed in Figure 2.4 (a) and (d). Then for the most modulated signals of Figure 2.3 (c) and (d), the smoothing effect of  $\lambda$  becomes less desirable as the noise level diminishes.

It is worth noting that, because the noise is strong, S-RD behaves better than MB-RD. Indeed, this can be explained by the fact that MB-RD relies on the unstable frequency modulation estimate  $\hat{q}_{\hat{f}}$ . In fact, the latter may point very far from the true IF of a mode, even if it is evaluated close to the IF, thus explaining why S-RD leads to better results.

In the sequel, we will study the improvement brought by RRP-RD for MR.

## 2.4. LINEAR CHIRP APPROXIMATION

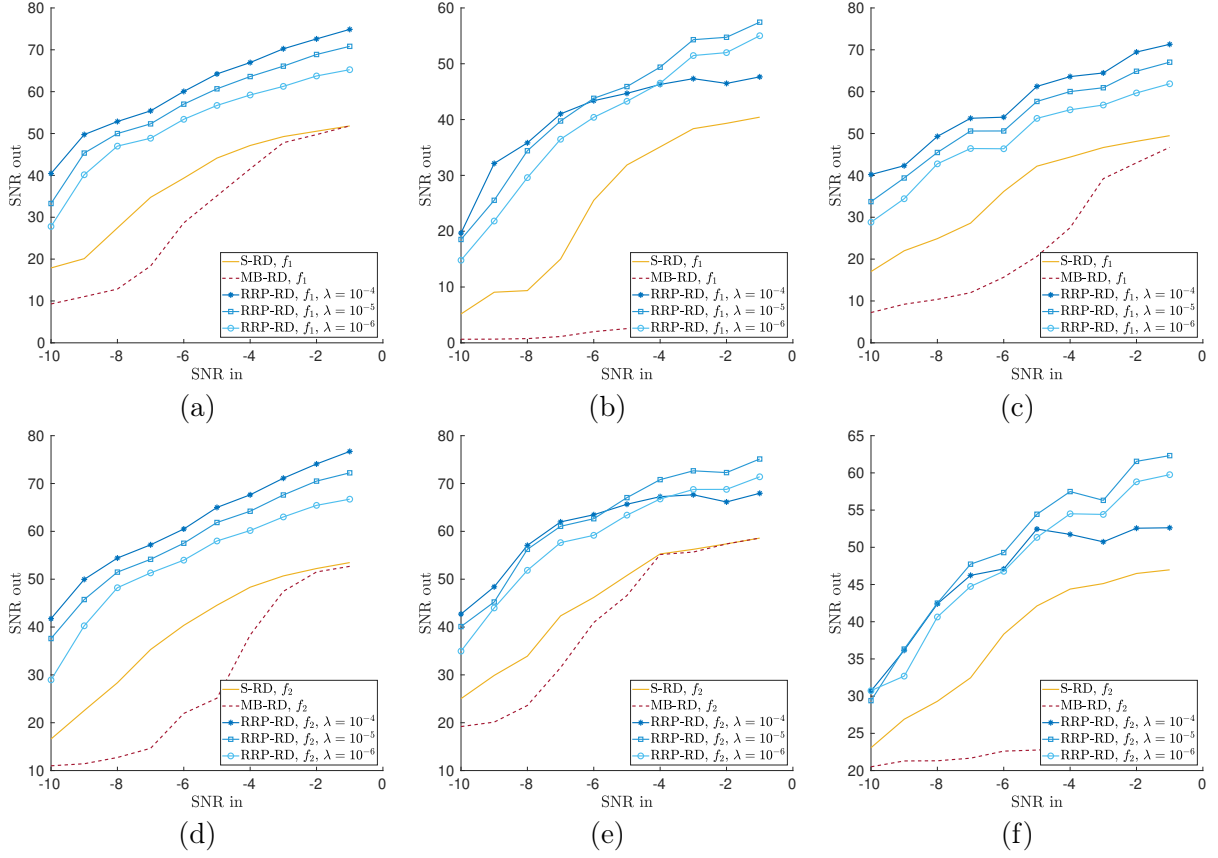


Figure 2.4.: (a): Comparison between S-RD, MB-RD and RRP-RD, for the mode  $f_1$  of the signal of Figure 2.3 (a), computation of output SNR between IF  $\phi'_1$  and estimated IF with respect to input SNR (the results are averaged over 40 realizations of noise); (b): same as (a) but for the mode  $f_1$  of the signal of Figure 2.3 (b); (c): same as (a) but for the mode  $f_1$  of the signal of Figure 2.3 (c); (d): same as (a) but the mode  $f_2$  of the signal of Figure 2.3 (a); (e): same as (b) but the mode  $f_2$  of the signal of Figure 2.3 (b); (f): same as (c) but the mode  $f_2$  of the signal of Figure 2.3 (c).

## 2.4. Linear Chirp Approximation

Once the ridges are detected, another issue is that of MR. For that purpose, we perform local TF analysis around the ridge to get an approximation of its associated mode, based on a simplified linear chirp model. Let us assume that the signal is a linear chirp of the form  $f(t) = Ae^{2i\pi(at+b\frac{t^2}{2})}$ , and  $g(t)$  is the Gaussian function  $e^{-\pi\frac{t^2}{\sigma^2}}$ . The STFT can then be seen as the Fourier transform of a Gaussian, and reads,

$$V_f^g(t, \eta) = f(t) \int_{\mathbb{R}} e^{-\pi\left(\frac{1}{\sigma^2} + i\phi''\right)(\tau-t)^2} e^{-2i\pi(\tau-t)(\eta-\phi'(t))} d\tau. \quad (2.22)$$



For a Gaussian function of the form  $h(t) = e^{-\pi z t^2}$ , where  $z = r e^{i\theta}$ , with  $r \in \mathbb{R}^+$  and  $\theta \in \mathbb{R}$ , such that  $\Re(z) > 0$ , its Fourier transform admits the following closed form,

$$\mathcal{F}(h)(\eta) = r^{-\frac{1}{2}} e^{-i\frac{\theta}{2}} e^{\frac{-\pi}{z} \eta^2}. \quad (2.23)$$

The proof is analogous to the case where  $z$  is real. Based on (2.23), one can deduce the analytic expression of  $V_f^g$  [19],

$$\begin{aligned} V_f^g(t, \eta) &= A r^{-\frac{1}{2}} e^{-i\frac{\theta}{2}} e^{-\frac{\pi \sigma^2 (1+ib\sigma^2)}{1+(b\sigma^2)^2} (\eta-a-bt)^2} \\ &= V_f^g(t, a+bt) e^{\frac{-\pi \sigma^2 (1+ib\sigma^2)}{1+(b\sigma^2)^2} (\eta-a-bt)^2} \\ &= V_f^g(t, \phi'(t)) e^{\frac{-\pi \sigma^2 (1+i\phi''(t)\sigma^2)}{1+(\phi''(t)\sigma^2)^2} (\eta-\phi'(t))^2}, \end{aligned} \quad (2.24)$$

with  $r = \sqrt{\frac{1}{\sigma^4} + \phi''(t)^2}$  and  $\theta = \tan^{-1}(-\phi''(t)\sigma^2)$ . If the monocomponent signal  $f$  departs from a linear chirp, the above equalities become approximations, which are furthermore only valid for  $\eta$  in the vicinity of  $\phi'(t)$ . Applying this approximation to  $f_p$ , we obtain for  $\eta$  in the vicinity of  $\phi'_p(t)$ ,

$$V_{f_p}^g(t, \eta) \approx V_{f_p}^g(t, \phi'_p(t)) e^{\frac{-\pi \sigma^2 (1+i\phi''_p(t)\sigma^2)}{1+(\phi''_p(t)\sigma^2)^2} (\eta-\phi'_p(t))^2} \quad (2.25)$$

Then, remarking that at TF location  $(t, \phi'_p(t))$ , the STFT of the mode  $f_q$ , for  $q \neq p$ , is close to 0 we replace in (2.25)  $V_{f_p}^g(t, \phi'_p(t))$  by  $V_f^g(t, \phi'_p(t))$ ,

$$V_{f_p}^g(t, \eta) \approx V_f^g(t, \phi'_p(t)) e^{\frac{-\pi \sigma^2 (1+i\phi''_p(t)\sigma^2)}{1+(\phi''_p(t)\sigma^2)^2} (\eta-\phi'_p(t))^2}. \quad (2.26)$$

This last equation gives us a way to approximate each mode in the general case of a MCS. However,  $\phi'_p(t)$ , and  $\phi''_p(t)$  are not known, and we are now going to explain how the FSST2 [19] provides reasonable values for  $\phi'_p(t)$  and  $\phi''_p(t)$ .

## 2.5. Approximation of $\phi'$ and $\phi''$ Based on FSST2 and Ridge Detection

The objective of this section is to explain how the IF and CR estimators involved in the definition of FSST2, introduced in Section 1.2, can be used in the model introduced

above. In fact, to get an acceptable level of accuracy, those estimates are to be evaluated in TF regions associated with high energy called ridges. Therefore, denoting the ridge by  $\Gamma_p$ , setting  $\psi'_p(t) = \widehat{\omega}_f^{[2]}(t, \Gamma_p(t))$ , and  $\psi''_p(t) = \widehat{q}_f(t, \Gamma_p(t))$ , this enables, from (2.24), to approximate the STFT of  $f_p$  by

$$V_{f_p}^g(t, \eta) \approx V_{\tilde{f}}^g(t, \psi'_p(t)) e^{-\frac{\pi\sigma^2(1+i\psi''_p(t)\sigma^2)}{1+(\psi''_p(t)\sigma^2)^2}(\eta - \psi'_p(t))^2}. \quad (2.27)$$

Finally, the reconstruction of mode  $f_p$  is carried out either through equation (1.5, page 2), (1.4, page 2), or (1.3, page 2) replacing  $V_f^g$  by the approximation of  $V_{f_p}^g$  given by (2.27).

It is important to remark here that the approximation of  $V_{f_p}^g$  uses only STFTs of  $\tilde{f}$  evaluated in the vicinity of the ridge associated with  $f_p$ , which will prove to be the very reason why the just described approach for mode retrieval is well adapted to handle the mode-mixing issue. However, up to now, we have only presented the approach for continuous time signals, and it needs to be adapted to discrete time signals, as explained hereafter.

## 2.6. Linear Chirp Based Mode Retrieval Technique

To implement the just described mode retrieval technique, we use the discrete setting of  $\tilde{f}$ , recalling that  $\tilde{f}[n] = \tilde{f}(\frac{n}{L})$ . In that context, the discrete mode  $(f_p[n])_n$  is associated with a ridge  $(\Gamma_p[n])_n$  corresponding to frequency indices in  $\{0, \dots, N-1\}$ . We then set  $\psi'_p[n] := \widehat{\omega}_f^{[2]}[n, \Gamma_p[n]]$  and  $\psi''_p[n] := \widehat{q}_f[n, \Gamma_p[n]]$  approximating  $\phi'_p(\frac{n}{L})$  and  $\phi''_p(\frac{n}{L})$  respectively. Remarking, using rectangular integration, that  $V_{\tilde{f}}^g(\frac{n}{L}, k\frac{L}{N}) \approx \frac{V_{\tilde{f}}^g[n, k]}{L}$ , using (2.26) and (2.27), we obtain,

$$\begin{aligned} V_{f_p}^g[n, k] &\approx LV_{f_p}^g\left(\frac{n}{L}, k\frac{L}{N}\right) \approx LV_{\tilde{f}}^g\left(\frac{n}{L}, \phi'_p\left(\frac{n}{L}\right)\right) e^{-\frac{\pi\sigma^2(1+i\phi''_p(\frac{n}{L})\sigma^2)}{1+(\phi''_p(\frac{n}{L})\sigma^2)^2}\left(\frac{kL}{N} - \phi'_p\left(\frac{n}{L}\right)\right)^2} \\ &\approx LV_{\tilde{f}}^g\left(\frac{n}{L}, \psi'_p[n]\right) e^{-\frac{\pi\sigma^2(1+i\psi''_p[n]\sigma^2)}{1+(\psi''_p[n]\sigma^2)^2}\left(\frac{kL}{N} - \psi'_p[n]\right)^2}, \end{aligned} \quad (2.28)$$

for which an approximation of  $V_{\tilde{f}}^g(\frac{n}{L}, \psi'_p[n])$  from values of the discrete STFT is needed. For that purpose, let us consider  $k_0 := \lfloor \psi'_p[n] \frac{N}{L} \rfloor$ , where  $\lfloor x \rfloor$  denotes the closest integer

to  $x$ . Because  $V_{\tilde{f}}^g(\frac{n}{L}, \psi'_p[n]) \approx V_{f_p}^g(\frac{n}{L}, \psi'_p[n])$ , using (2.27), we may write,

$$\begin{aligned} V_{\tilde{f}}^g\left(\frac{n}{L}, \psi'_p[n]\right) &\approx V_{\tilde{f}}^g\left(\frac{n}{L}, k_0 \frac{L}{N}\right) e^{\frac{\pi \sigma^2 (1+i\psi_p''[n]\sigma^2)}{1+(\psi_p''[n]\sigma^2)^2} (k_0 \frac{L}{N} - \psi'_p[n])^2} \\ &\approx \frac{1}{L} V_{\tilde{f}}^g[n, k_0] e^{\frac{\pi \sigma^2 (1+i\psi_p''[n]\sigma^2)}{1+(\psi_p''[n]\sigma^2)^2} (k_0 \frac{L}{N} - \psi'_p[n])^2}. \end{aligned} \quad (2.29)$$

From (2.28) and (2.29) we get that,

$$\begin{aligned} V_{f_p}^g[n, k] &\approx V_{\tilde{f}}^g[n, k_0] e^{\frac{\pi \sigma^2 (1+i\psi_p''[n]\sigma^2)}{1+(\psi_p''[n]\sigma^2)^2} \left[ \left(\frac{k_0 L}{N} - \psi'_p[n]\right)^2 - \left(\frac{k L}{N} - \psi'_p[n]\right)^2 \right]} \\ &\approx V_{\tilde{f}}^g[n, k_0] e^{\frac{\pi \sigma^2 (1+i\psi_p''[n]\sigma^2)}{1+(\psi_p''[n]\sigma^2)^2} \left[ \frac{L(k_0 - k)}{N} \left(\frac{L(k_0 + k)}{N} - 2\psi'_p[n]\right) \right]}. \end{aligned} \quad (2.30)$$

If we denote  $\tilde{V}_{f_p}^g$  the estimation of  $V_{f_p}^g$  given by (2.30), the retrieval of  $f_p$  can be carried out either through (1.7, page 3), if  $g[0] \neq 0$ ,

$$f_p[n] \approx \frac{1}{g[0]N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \tilde{V}_{f_p}^g[n, k], \quad (2.31)$$

or assuming  $f$  is  $L$ -periodic, through (1.8, page 3), if we assume  $f$  is  $L$ -periodic,  $g \neq 0$ ,

$$f_p[n] \approx \frac{1}{N\|g\|_2} \sum_{k=0}^{N-1} \sum_{m=n-M}^{n+M} \tilde{V}_{f_p}^g[n, k] g[n-m] e^{i2\pi \frac{k(n-m)}{N}}. \quad (2.32)$$

In a noisy context, all the above expressions can be used replacing  $f$  by its noisy version  $\tilde{f}$ . The techniques given by (2.31) and (2.32) are respectively denoted  $M_1$  and  $M_2$ . When making abstraction of the reconstruction given by (1.7, page 3) and (1.8, page 3), this technique is called *linear chirp based retrieval* (LCR).

## 2.7. Alternative Techniques for Mode Retrieval: HT, SSR-HT

The HT technique when used for the retrieval of the mode  $f_p$  of a MCS [15] consists of considering in the reconstruction process only some coefficients of  $V_{\tilde{f}}^g$  in the vicinity of  $\tilde{\Gamma}_p$ , the ridge associated with  $\tilde{f}_p$  the noisy version of  $f_p$ . For that purpose, one defines

for each time indexed by  $n$ , an interval  $J_p[n] = [\eta_p^-[n], \eta_p^+[n]]$  such that:

$$\begin{aligned}\eta_p^-[n] &:= \operatorname{argmax}_k \left\{ \frac{kL}{N} < \tilde{\Gamma}_p[n], |V_{\tilde{f}}^g[n, k]| < 3\hat{\gamma} \right\} \\ \eta_p^+[n] &:= \operatorname{argmin}_k \left\{ \frac{kL}{N} > \tilde{\Gamma}_p[n], |V_{\tilde{f}}^g[n, k]| < 3\hat{\gamma} \right\}.\end{aligned}\quad (2.33)$$

The reconstruction of the modes is carried out summing the coefficients over  $J_p$  using (1.7, page 3), if  $g[0] \neq 0$ ,

$$f_p[n] \approx \frac{1}{g[0]N} \sum_{k \in J_p[n]} V_{\tilde{f}}^g[n, k], \quad (2.34)$$

or assuming  $f$  is  $L$ -periodic, through (1.8, page 3), if we assume  $f$  is  $L$ -periodic,  $g \neq 0$ ,

$$f_p[n] \approx \frac{1}{N\|g\|_2} \sum_{k \in J_p[n]} \sum_{m=n-M}^{n+M} V_{\tilde{f}}^g[n, k] g[n-m] e^{i2\pi \frac{k(n-m)}{N}}. \quad (2.35)$$

Similarly to LCR, we respectively name those techniques HT  $M_1$  and HT  $M_2$ .

For the sake of comparison, we also briefly recall the principles of a denoising technique called *shifted-symmetrized-regularized hard-thresholding* (SSR-HT), and introduced in [58], which consists of an improvement of HT technique based on a linear chirp approximation for the mode. It first shifts the real and imaginary parts of  $V_{\tilde{f}}^g$  such that their respective maximum in  $J_p$  is centered on  $\Gamma_p$  in the frequency axis. We denote the shifted STFT by  $VS_{p,\tilde{f}}^g$ . Then the symmetrization is defined by,

$$SYM_{\tilde{f},p}^g[n, k] = \frac{1}{2} \left( VS_{p,\tilde{f}}^g[n, k] + VS_{p,\tilde{f}}^g[n, 2\Gamma_p[n] - k] \right) \times \mathbb{1}_{J_p}[k]. \quad (2.36)$$

where  $\mathbb{1}_{J_p}[k] = 1$  if  $k \in J_p$  and  $\mathbb{1}_{J_p}[k] = 0$  otherwise. Next, SSR-HT smoothes real and imaginary parts of  $SYM_{\tilde{f},p}^g$  using piecewise cubic Hermite interpolation, that we denote by  $SSR_{\tilde{f},p}^g$ . Finally, as for LCR and HT, if  $g[0] \neq 0$ ,

$$f_p[n] \approx \frac{1}{g[0]N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} SSR_{\tilde{f},p}^g[n, k], \quad (2.37)$$

or assuming  $f$  is  $L$ -periodic, through (1.8, page 3), if we assume  $f$  is  $L$ -periodic,  $g \neq 0$ ,

$$f_p[n] \approx \frac{1}{N\|g\|_2} \sum_{k=0}^{N-1} \sum_{m=n-M}^{n+M} SSR_{\tilde{f},p}^g[n, k] g[n-m] e^{i2\pi \frac{k(n-m)}{N}}. \quad (2.38)$$

Here again, similarly to LCR and HT, we respectively denote the two reconstruction techniques by SSR  $M_1$  and SSR  $M_2$ .

## 2.8. Numerical Applications

In this section, we first show the benefits of LCR when the input SNR is between 0 and 10 dB. At such noise levels, all RD are working fine and RRP-RD does not bring any significant improvement. Therefore, we assume that there are no problems with RD, and we focus on comparing HT, SSR-HT with LCR. Then, we further study HT and LCR techniques, when the input SNR is between  $-10$  and 0 dB, in which cases, RRP-RD is used to detect the ridges. Finally, we show that LCR can be used to solve the mode mixing issue.

### 2.8.1. Comparison of Mode Retrieval Techniques on a Monocomponent Signal

We study the efficiency of LCR for the signals of Figure 2.5 (a), (b) and (c); the results being depicted in Figure 2.6. As the input SNR varies between 0 and 10 dB, we use S-RD as it is working fine at such noise levels.

We notice that LCR gives better output SNR than HT and SSR-HT in each case, whatever the noise level. We also note that LCR is only slightly sensitive to the technique used for mode reconstruction ( $M_2$  seems a little better than  $M_1$ , but the latter is much faster than the former). On the contrary, SSR-HT behaves worse when the modulation decreases, since the quality of reconstruction is hampered by frequency resolution when  $M_2$  is used. Finally, comparing the results of Figure 2.6 (b) and (c), we notice that LCR is efficient for a wide range of frequency modulation for the mode.

### 2.8.2. Study of the Robustness of Mode Retrieval

In this section, we investigate the robustness of mode retrieval techniques *MR based on S-RD and LCR* (S-LCR-MR) and *MR based on RRP-RD and LCR* (RRP-LCR-MR), and draw a conclusion of the impact of ridge detection on MR.

First, RRP-LCR-MR provides significant improvement compared with other MR techniques that rely on RRP-RD [40]. Now, we focus on the impact of RD in heavy noise situations. To do so, we consider the signals of Figure 2.3 (a), (b), (c), for which we obtain the results of Figure 2.7 (a), (b), (c) respectively. In these figures, RRP-LCR-MR is compared with S-LCR-MR and the benefits of using RRP-RD transpires in the better

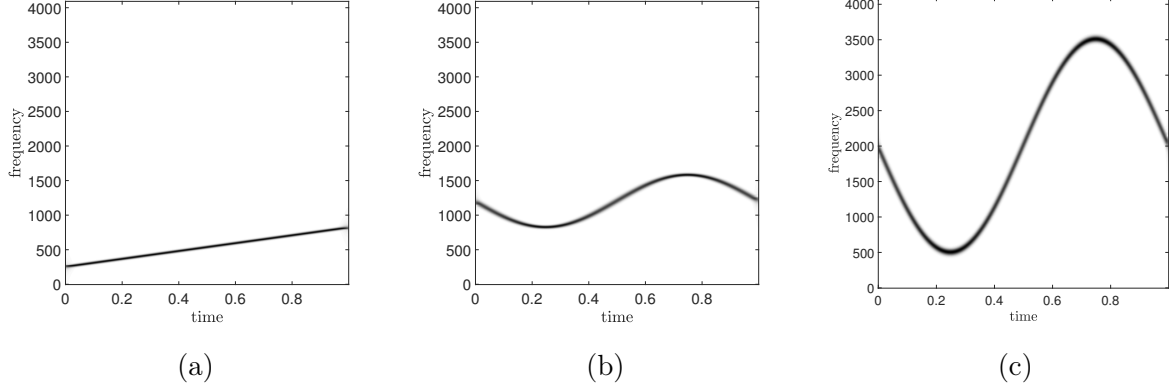


Figure 2.5.: (a): STFT of  $f(t) = e^{2i\pi(250t+568\frac{t^2}{2})}$ ; (b): STFT of  $f(t) = e^{2i\pi(1200t+60\cos(2\pi t))}$ ; (c): STFT of  $f(t) = e^{2i\pi(2000t+238\cos(2\pi t))}$ .

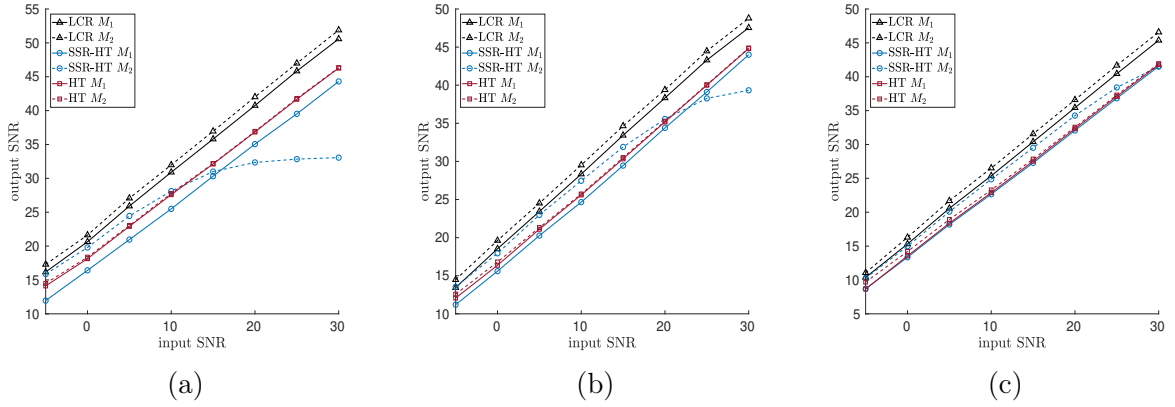


Figure 2.6.: (a): Signal reconstruction results for the signal of Figure 2.5 (a) using HT, SSR-HT and LCR on the ridge provided by S-RD. the mode is reconstructed using either (1.7, page 3) or (1.8, page 3) ; (b): Same as (a) but for the signal depicted in Figure 2.5 (b); (c): Same as (a) but for the signal depicted in Figure 2.5 (c).

output SNR. Even when the noise has a strong impact on the entire STFT, a good ridge detection is necessary to improve MR. In fact, the accuracy of RRP-RD allows to get more accurate IF and CR estimates, by using  $s_p^{fin}$  instead of  $\hat{\omega}_{\hat{f}}^{[2]}$  and  $(s_p^{fin})'$  instead of  $\hat{q}_{\hat{f}}$ . This can be checked numerically, as depicted in Figure 2.8 and justify their usage as IF and CR estimators.

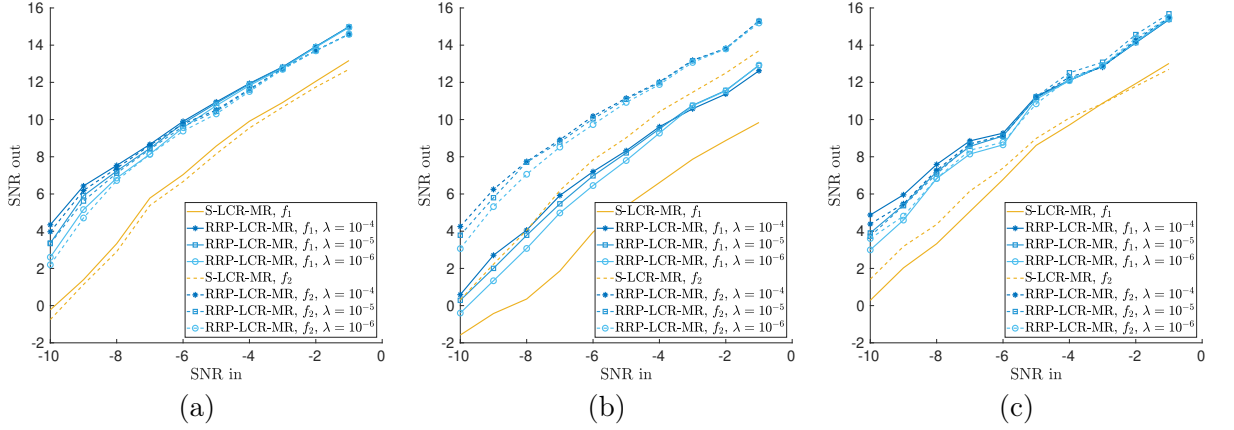


Figure 2.7.: (a): For each mode  $p = 1, 2$ , output SNR between mode  $f_p$  of signal of Figure 2.3 (a) and reconstructed mode using either S-LCR-MR or RRP-LCR-MR (the results are averaged over 40 noise realizations); (b): same but with signal of Figure 2.3 (b); (c): same but with signal of Figure 2.3 (c);

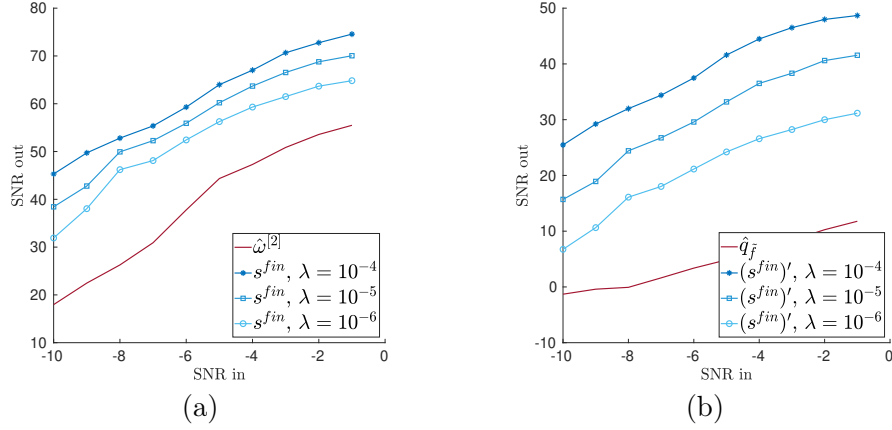


Figure 2.8.: (a): computation of the output SNR associated with the estimation of  $\phi'$  with  $s^{fin}$  (computed for different values of  $\lambda$ ) or with  $\hat{\omega}^{[2]}$  for the first mode of Figure 2.3 (a); (b): computation of the SNR associated with the estimation of  $\phi''$  with  $(s^{fin})'$  (computed for different values of  $\lambda$ ) or with  $\hat{q}_{\hat{f}}$  for the first mode of Figure 2.3 (a).

### 2.8.3. Mode Mixing Issue

In this section, we study the ability offered by HT, SSR-HT and LCR to accurately separate two close modes in a MCS provided they admit two separate ridges. When the modes are close in the TF plane, the modulus of the STFT between the two modes is oscillating, due to interference. Therefore, some coefficients located between two modes cannot be reasonably associated with one of them. This phenomenon is often called mode mixing in the literature. It was studied in the context MR, by comparing *empirical*

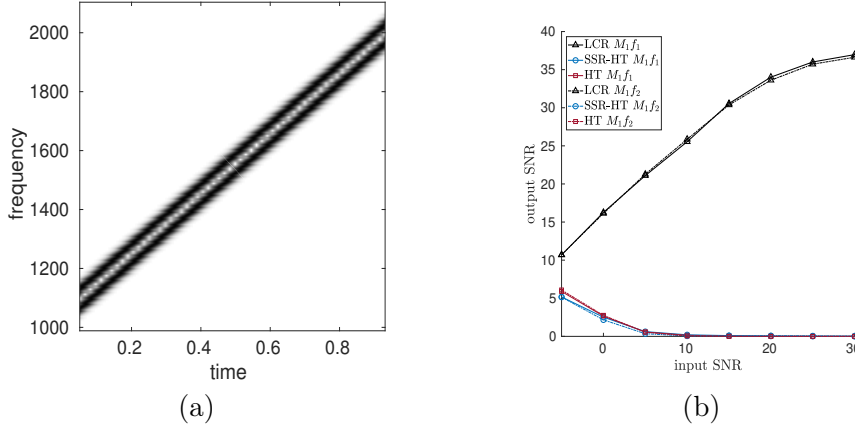


Figure 2.9.: (a): STFT of signal made of two close chirps; (b): Reconstruction results using (1.7, page 3) for each of the two modes depicted in (a) using S-RD.

*mode decomposition* and FSST [18], and also in [59], to improve the *empirical mode decomposition*. In such a case, one may also consider the adaptive transforms [36], and perform MR based on the *signal separation operator*.

To illustrate how we tackle this issue, we consider the two mode signal whose STFT is depicted in Figure 2.9 (a), for which mode separation is challenging. In particular, we notice that at a low noise level and when using HT or SSR-HT, the threshold used for the determination of intervals  $J_p$  is lower, at many time instants, than the minimum value of the modulus of STFT between the two ridges associated with the modes. This results in mode mixing with these two techniques. In Figure 2.9 (b), we display the results corresponding to the retrieval of the modes  $f_1$  and  $f_2$ . With HT and SSR-HT techniques, when the noise level is low, we get that for many time indices  $n$ ,  $J_1[n] = J_2[n]$ , hence the poor results. On the contrary, the separation of the two modes with LCR remains satisfactory whatever the noise level.

As a remark, the reconstruction formula used with HT, SSR-HT, and LCR is (1.7, page 3), to use (1.8, page 3) instead would not change the conclusions.

## 2.9. Application to Gravitational-Wave Signals

In this section, we investigate the applicability of RRP-RD and RRP-LCR-MR to a transient gravitational-wave signal, generated by the coalescence of two stellar-mass black holes. This event, called **GW150914**, was detected by the LIGO detector Hanford, Washington and closely matches the waveform Albert Einstein predicted almost 100 years ago in his general relativity theory for the inspiral, the merger of a pair of black



holes and the ringdown of the resulting single black hole [60]. The observed signal has a length of 3441 samples in  $T = 0.21$  seconds.

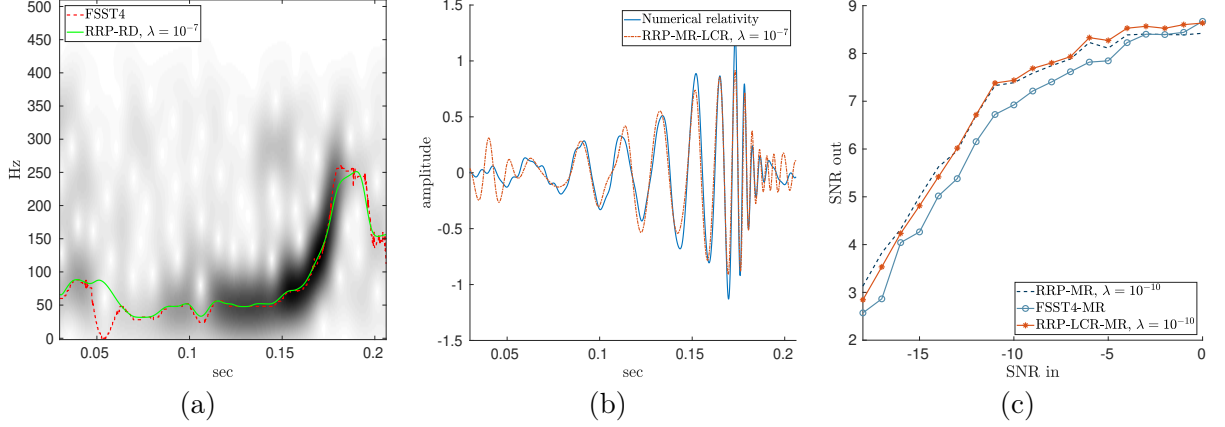


Figure 2.10.: (a): STFT modulus ( $\sigma = 0.0105$ ) of the Hanford signal along with the ridge obtained with RRP-RD and FSST4; (b): illustration of signal reconstruction based on RRP-LCR-MR and the numerical relativity; (c): SNR corresponding to the reconstruction of the signal using either RRP-LCR-MR, RRP-MR or FSST4-MR (the ground truth is assumed to be the one produced by numerical relativity). The results are averaged over 40 realizations of noise.

Because the signal is real, it cannot be analytic and we therefore use the reconstruction formula associated with real signals,

$$\tilde{f}[n] = \frac{2}{g[0]N} \Re \left\{ \sum_{k=0}^{\frac{N}{2}-1} V_f^g[n, k] \right\}. \quad (2.39)$$

For the sake of comparison, we consider the *fourth-order Fourier-based synchrosqueezing transform* (FSST4), an efficient reassignment technique introduced in [61]. Note that, as mentioned in Section 1.2, the generalized FSSTN technique will be defined in Chapter 4. We first display in Figure 2.10 (a), the modulus of the STFT of such a signal, along with the ridge associated with FSST4 and the spline obtained using RRP-RD. We notice that RRP-RD and FSST4 leads to very similar results, and that both techniques enable the detection of the ringdown.

We then investigate the quality of signal reconstruction by comparing it with the one given by the numerical relativity [62], when one uses either *MR based on RRP-RD and HT* (RRP-MR), RRP-LCR-MR or *MR based on FSST4* (FSST4-MR). An illustration of the reconstructed signal obtained with RRP-LCR-MR superimposed on the one given by the numerical relativity is displayed in Figure 2.10 (b). Then, we

estimate the quality of signal reconstruction with the three techniques when the input SNR varies. The results depicted in Figure 2.10 (c) show that the three methods behave similarly, though RRP-MR and RRP-LCR-MR are always slightly better. What is very specific to the studied gravitational wave signal is that the part associated with the strong frequency modulation is very energetic, and a slight inaccuracy in IF estimation using the spline approximation at the locations corresponding to strong frequency modulation has a strong impact on mode reconstruction with RRP-LCR-MR. For this reason, it may not be that interesting to use the latter technique instead of RRP-MR in that context. It is also worth noting here that FSST4-MR is very specific to signals containing very oscillatory phases, which is the case of the gravitational wave when the ringdown occurs. For any other modes like those studied before, a lower order synchrosqueezing transform would lead to better results [63]. Another limitation of the techniques based on synchrosqueezing transforms is that they are not adapted to long signals since the reconstruction of the modes from these transforms when the hop-size is larger than one is not tractable [15]. Finally, as FSST4-MR uses S-RD, it depends on the user defined parameter  $B_f$  which is not the case of RRP-MR or RRP-LCR-MR. These last techniques thus offer a nice alternative to the best state-of-the-art techniques for the reconstruction of very complex signals such as gravitational-wave signals.

## 2.10. Conclusion

In this chapter, we proposed a new ridge detection technique and investigated its potential use for MR. We introduced the combination of the so-called RRP-RD and LCR techniques, which leads to satisfactory MR results. However, when considering heavy noise, it becomes a lot more challenging since the noise affects the coefficients of the entire TF plane. In particular, we showed that even with very accurate IF and CR estimates, MR is still challenging and the proposed combination of RRP-RD and LCR is a good answer to that problem. However, it is important to note that our approach is unable to deal with crossing modes, since, in that case, one cannot associate a ridge with each mode. This requires a better understanding of interference, which will be the topic of Chapter 3. We will also see in Chapter 4, that the noise still has a huge impact on the amplitude of the modes and, therefore an ideal RD is not sufficient for accurate MR.



## 3. Time Frequency Analysis of Interference

### 3.1. Introduction

In this chapter, we study the time and frequency interference of a MCS, which arise when two modes are close in the TF plane, thus creating interference patterns.

In a first part of this chapter, we carry out the mathematical study of the conditions for interference patterns called TFBs to appear. TFBs were originally introduced in [64] with mathematical characterizations of this phenomenon in the case of two pure harmonics. We also study this situation and extend it to signals made of two parallel linear chirps. Then, because in practice, the conditions for TFBs to appear are difficult to evaluate, and for the purpose of considering general MCSs, we detail a special RD that is able to detect the merging of ridges that coincides with the presence of TFBs. We then perform a numerical validation of the technique, on noisy MCSs.

In the second part of this chapter, we consider a different approach to detect interference patterns based on the analysis of the zeros of the spectrogram, and show perspectives on how to perform a separation between different types of zeros.

### 3.2. On the Existence of Time Frequency Bubbles

In this section, we investigate the number of existing *local maxima of the spectrogram along the frequency axis* (LMFs) when two modes are interfering. Depending on the situation in time and frequency, we show that when modes are merging or are too close in the TF plane, there may only be one LMF associated with both modes.

In what follows, to characterize these situations, we will use the notion of TFB, which corresponds to the following definition:

---

This chapter is based on the work by S. Meignen, T. Oberlin, and N. Laurent, “On the Study of Interference and their Localization in the Time-Frequency Plane” submitted to IEEE TSP.

**Definition 3.2.1.** *Two modes make a TFB in the TF plane when the LMFs associated with each mode merge.*

We now investigate the existence of TFBs when the signal is the sum of two pure tones and then explain how the obtained results can be extended to the case of two parallel linear chirps. As explained in Chapter 2, a Gaussian window is well adapted for such a study as it allows working with closed form expressions thus enabling the mathematical analysis. Therefore, we set the window  $g(t) = e^{-\pi \frac{t^2}{\sigma^2}}$ .

### 3.2.1. Case of the Sum of Two Pure Tones

Let  $f(t) = f_1(t) + f_2(t)$  with  $f_1(t) = Ae^{i2\pi\xi_1 t}$  and  $f_2(t) = e^{i2\pi\xi_2 t}$ , where  $\xi_1 < \xi_2$ . As one has  $V_{f_1}^g(t, \eta) = \sigma Ae^{i2\pi\xi_1 t} e^{-\pi(\eta-\xi_1)^2 \sigma^2}$  and  $V_{f_2}^g(t, \eta) = \sigma e^{i2\pi\xi_2 t} e^{-\pi(\eta-\xi_2)^2 \sigma^2}$ ,

$$|V_f^g(t, \eta)|^2 = \sigma^2 \left( A^2 e^{-2\pi\sigma^2(\eta-\xi_1)^2} + e^{-2\pi\sigma^2(\eta-\xi_2)^2} + 2Ae^{-\pi\sigma^2[(\eta-\xi_1)^2 + (\eta-\xi_2)^2]} \cos(2\pi(\xi_2 - \xi_1)t) \right). \quad (3.1)$$

A first property of  $|V_f^g(., \eta)|^2$ , where  $.$  means we consider this variable, is that it attains its maximum at  $t_k = \frac{k}{\xi_2 - \xi_1}$ ,  $k \in \mathbb{Z}$  for which:

$$|V_f^g(t_k, \eta)|^2 = \sigma^2 (Ae^{-\pi\sigma^2(\eta-\xi_1)^2} + e^{-\pi\sigma^2(\eta-\xi_2)^2})^2. \quad (3.2)$$

Denoting the *local extrema of the spectrogram along the frequency axis* by *local extrema of the spectrogram along the frequency axis* (LEFs), the study of  $|V_f^g(t_k, .)|^2$  leads to:

**Proposition 3.2.1.**  $|V_f^g(t_k, .)|^2$  has three LEFs (two LMFs and a local minimum) if and only if  $\alpha := \sqrt{\frac{\pi}{2}}\sigma(\xi_2 - \xi_1) > 1$  and

$$|\log(A)| < -2 \operatorname{arccosh}(\alpha) + 2\alpha\sqrt{\alpha^2 - 1}.$$

In any other case,  $|V_f^g(t_k, .)|^2$  has a unique extremum which is a LMF.

The proof is given in Appendix A.1. Having studied the situation at  $t_k$ , we can generalize this result to any time  $t$  through:

**Proposition 3.2.2.**  $|V_f^g(t, .)|^2$  has three LEFs (two LMFs and a local minimum) if and only if  $\alpha := \sqrt{\frac{\pi}{2}}\sigma(\xi_2 - \xi_1) > 1$  and

$$|\log(A)| < -2 \operatorname{arccosh}(\alpha) + 2\alpha\sqrt{\alpha^2 - 1}.$$

In any other case, there exists some time  $t$  where  $|V_f^g(t, \cdot)|^2$  has a unique extremum which is a LMF.

The proof is given in Appendix A.2. This proposition means that, if the spectrogram exhibits two LMFs at  $t_k$ , it is the same for any time  $t$ , and it tells us when two pure tones are generating TFBs.

Note that, when  $A = 1$  and since  $-2 \operatorname{arccosh}(\alpha) + 2\alpha\sqrt{\alpha^2 - 1}$  is an increasing function which equals 0 at 1, when  $\alpha > 1$  we always have  $-2 \operatorname{arccosh}(\alpha) + 2\alpha\sqrt{\alpha^2 - 1} > 0$ , and thus the second condition in Proposition 3.2.2 is useless. As a consequence, when  $A = 1$  TFBs exist only when  $\alpha \leq 1$ . Also, comparing the hypotheses in Propositions 3.2.1 and 3.2.2, we notice that the presence of TFBs is only determined by the behavior of the spectrogram at time  $t_k$ . Finally, denoting by  $(t_k, \eta_k)$  the location of a LMF at  $t_k$ , this TF point also corresponds to a maximum of the spectrogram along the time axis and thus it is a local maximum of the bidimensional function  $|V_f^g(t, \eta)|^2$ .

To give a first illustration of TFBs we consider the spectrogram of the sum of two pure tones, in Figure 3.1 (a) and (b),  $A = 1$  and the conditions of Proposition 3.2.2 are fulfilled in (a) but not in (b) TFBs being thus present in the latter case. Finally, in Figure 3.1 (c),  $A \neq 1$  and the conditions of Proposition 3.2.2 are fulfilled. In each case, we also plot the zeros and the local maxima of the spectrogram along with the spectrogram ridges corresponding to the chaining along the time axis of LMFs. As expected, the local maxima are located on spectrogram ridges at  $t = t_k$ .

### 3.2.2. Case of Two Parallel Linear Chirps

Our goal is now to extend the previous study to parallel linear chirps. We thus consider  $f(t) = f_1(t) + f_2(t)$  with  $f_1(t) = Ae^{2i\pi\phi_1(t)}$  and  $f_2(t) = e^{2i\pi\phi_2(t)}$ , with  $\phi_k(t) = a_k t + \frac{C}{2}t^2$  with  $C$  constant.

As one has a similar formula to (2.24, page 24),

$$V_{f_1}^g(t, \eta) = f_1(t)r^{-1/2}e^{-\frac{\pi\sigma^2(1+iC\sigma^2)(\eta-\phi_1'(t))^2}{1+C^2\sigma^4}}e^{-i\frac{\theta}{2}} \quad (3.3)$$

with  $r = \frac{\sqrt{1+C^2\sigma^4}}{\sigma^2}$  and  $\theta = \operatorname{atan}(-C\sigma^2)$ , one obtains:

$$\begin{aligned} |V_f^g(t, \eta)|^2 = & \frac{1}{r} \left( A^2 e^{-2\pi \frac{(\eta-\phi_1'(t))^2}{r^2\sigma^2}} + e^{-2\pi \frac{(\eta-\phi_2'(t))^2}{r^2\sigma^2}} \right. \\ & \left. + 2Ae^{-\pi \frac{[(\eta-\phi_1'(t))^2 + (\eta-\phi_2'(t))^2]}{r^2\sigma^2}} \cos(2\pi\phi(t, \eta)) \right), \end{aligned} \quad (3.4)$$

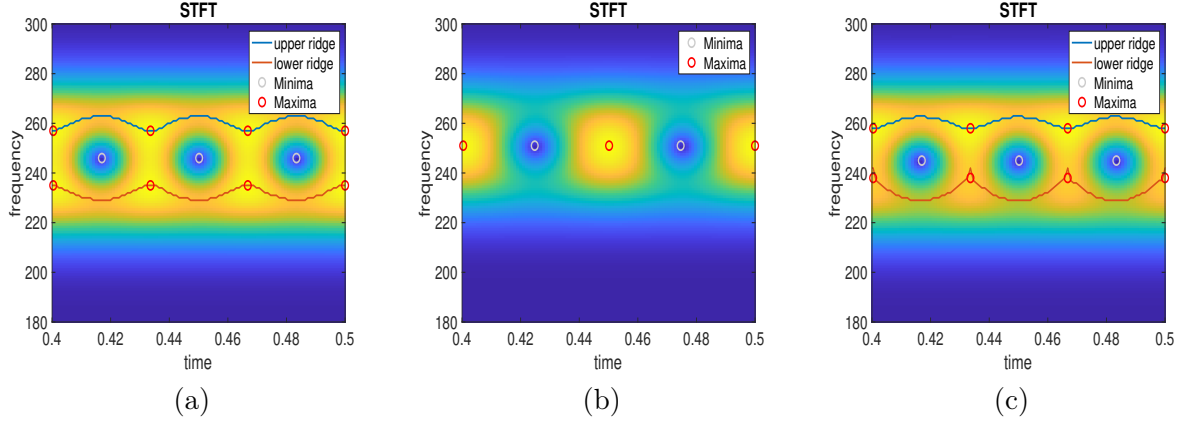


Figure 3.1.: (a): spectrogram of two parallel pure harmonic modes with the same amplitude, when there exists three LEFs at each time instant (the ridges corresponding to the chaining of LMFs are also superimposed); (b): spectrogram of two parallel pure harmonic modes exhibiting TFBs; (c): same as (a) but when the pure harmonic modes have different amplitude (three LEFs are still present at each time instant but the symmetry no longer exists). In each case, we also plot the zeros and the local maxima of the spectrogram.

with  $\phi(t, \eta) = (a_2 - a_1) \left( t + \frac{C}{r^2 \sigma^2} \left( \eta - Ct - \frac{a_1 + a_2}{2} \right) \right)$ . Let us then consider the set of TF points  $(t, \eta)$  such that  $\cos(2\pi\phi(t, \eta)) = \lambda$ . These points correspond to

$$\phi(t, \eta) = \frac{\arccos(\lambda)}{2\pi} + k \text{ for } k \in \mathbb{Z} \Leftrightarrow t = D\eta + B_k^\lambda \quad (3.5)$$

with  $D = \frac{C}{C^2 - r^2 \sigma^2}$  and

$$B_k^\lambda = \frac{C^2(a_2^2 - a_1^2) + r^2 \sigma^2 \left( \frac{\arccos(\lambda)}{\pi} + 2k \right)}{2(r^2 \sigma^2 - C^2)(a_2 - a_1)}.$$

We deduce from (3.4) that:

$$|V_f^g(D\eta + B_k^1, \eta)|^2 = \frac{1}{r} \left( A e^{-\pi \left( \frac{1-CD}{r\sigma} \right)^2 \left( \eta - \frac{a_1 + CB_k^1}{1-CD} \right)^2} + e^{-\pi \left( \frac{1-CD}{r\sigma} \right)^2 \left( \eta - \frac{a_2 + CB_k^1}{1-CD} \right)^2} \right)^2, \quad (3.6)$$

which is the same expression as (3.2), replacing  $\sigma$ ,  $\xi_1$  and  $\xi_2$  by  $\frac{1-CD}{r\sigma}$ ,  $\frac{a_1 + CB_k^1}{1-CD}$  and  $\frac{a_2 + CB_k^1}{1-CD}$ , and thus one can deduce the following

**Proposition 3.2.3.**  $|V_f^g(t, \eta)|^2$  has three extrema (two maxima and one minimum) along

the line  $t = D\eta + B_k^1$  if and only if  $\alpha_C = \sqrt{\frac{\pi}{2} \frac{a_2 - a_1}{r\sigma}} > 1$ , and then

$$|\log(A)| < -2 \operatorname{arccosh}(\alpha_C) + 2\alpha_C \sqrt{\alpha_C^2 - 1},$$

and a unique maximum along that line otherwise.

The proof is the same as that of Proposition 3.2.1. Then one generalizes Proposition 3.2.2 as follows:

**Proposition 3.2.4.**  $|V_f^g(t, \eta)|^2$  has three extrema (two maxima and one minimum) along the line  $t = D\eta + B_k^\lambda$  if and only if  $\alpha_C = \sqrt{\frac{\pi}{2} \frac{a_2 - a_1}{r\sigma}} > 1$ , and then

$$|\log(A)| < -2 \operatorname{arccosh}(\alpha_C) + 2\alpha_C \sqrt{\alpha_C^2 - 1},$$

and a unique maximum along that line otherwise.

So the presence of TFBs is here related to what happens when  $\lambda = 1$ . Let us denote by  $(t_k, \eta_k)$  a local maximum on the line  $t = D\eta + B_k^1$  and then a neighboring point  $(t, \eta)$  of  $(t_k, \eta_k)$  in the direction of vector  $(1, C)$ . For such a point, one has  $\eta - \phi_1'(t) = \eta_k - \phi_1'(t_k)$  and  $\eta - \phi_2'(t) = \eta_k - \phi_2'(t_k)$  and, as  $\cos(2\pi\phi(t, \eta)) < 1$ , one gets that  $(t_k, \eta_k)$  is also a maximum in the direction of vector  $(1, C)$ , and thus a local maximum of the bidimensional function  $|V_f^g(t, \eta)|^2$ . Note finally that, contrary to what happens in the pure tones case,  $(1, C)$  is not orthogonal to  $(D, 1)$ .

An illustration is given in Figure 3.2 (a) where the spectrogram of two parallel linear chirps is displayed and for which there exist three extrema along direction  $t = D\eta + B_k^\lambda$ . Then, in Figure 3.2 (b), the case where there exists only one maximum for some value of  $\lambda$  is displayed (TFBs are thus present). Finally, in Figure 3.2 (c) one considers a case where there are three extrema along direction  $t = D\eta + B_k^\lambda$  when  $A$  is no longer equal to 1. In each case, the zeros and the local maxima of the spectrogram are also plotted.



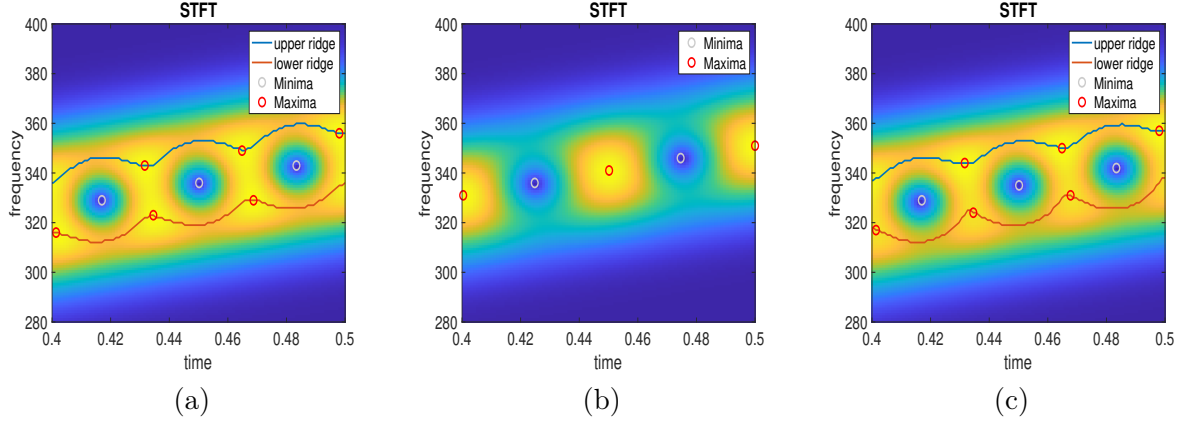


Figure 3.2.: (a): Spectrogram of two parallel linear chirps with the same amplitude, when there always exists three extrema in the direction  $D$ ; (b): Spectrogram of two parallel linear chirps when TFBs are present. Three extrema no longer exist in the direction  $D$ ; (c): same as (a) but when the linear chirps have different amplitude (three extrema are still present along the direction  $D$ ). In (a) to (c), we also plot the zeros and the local maxima of the spectrogram.

### 3.3. Localizing TFBs Based on a New RD

In this section, we design a technique for the detection of TFBs. Indeed, the conditions we expressed above require the knowledge on the frequency and modulation of the signal, which is not the case in practice.

From the definition of TFBs, we propose to define a new RD to find the LMFs that are merging. However, the previously introduced RD such as S-RD in Section 2.2.1, MB-RD in Section 2.2.2 and RRP-RD in Section 2.3.3 are not compatible for such a task, because they do not allow ridges to cross. Therefore, in the section below, we explain how to build a new RD enabling mode crossing.

#### 3.3.1. New RD Enabling the Merging of Ridges

In Chapter 2, the RDs we introduced are specifically designed to work for non-crossing modes and cannot be used as is to detect TFBs. To build the new RD enabling the merging of ridges, and because we only consider input SNR above 5dB, we propose to modify S-RD introduced in Section 2.2.1, which gives good results in such situations. In the case of S-RD, the ridge  $\Gamma_p$  defined by means of (2.6, page 13) cannot cross the previously detected ridges, and is not necessarily made of LMFs. To circumvent these two limitations of S-RD, we still use the same initialization process, namely, for an

initialization at time index  $n_0$ ,

$$\Gamma_p[n_0] = \arg \max_{0 \leq k \leq N-1} |V_{\tilde{f}}^g[n_0, k]|, \quad (3.7)$$

but replace (2.6, page 13) by

$$\Gamma_p[n+1] := \operatorname{argmin}_k \left\{ |k - \Gamma_p[n]| \leq B_f, \text{ s.t. } |V_{\tilde{f}}^g[n+1, k]| \text{ LMF} \right\}. \quad (3.8)$$

This has two straightforward consequences. First, it is possible that  $\Gamma_p$  merges with some of the previously computed ridges, and second, the ridge points are all LMFs. As a consequence, if no LMF is found in the interval of interest, the construction of the ridge is stopped, which was not the case with S-RD. The second consequence is that, contrary to S-RD, when two LMFs are found in the interval of interest one considers the closest to the current position rather than the global maximum on the interval.

Note that to extract a ridge with S-RD, the procedure is run with different initializations and the ridge associated with the largest energy is kept. When using (3.8) to compute  $\Gamma_p$  instead, the latter may have points in common with the previously computed ridges, thus the energy we associate with  $\Gamma_p$  corresponds this time to the sum of the spectrogram coefficients that do not belong to the previously computed ridges. Meanwhile we also define the length of  $\Gamma_p$  as the length of the ridge portions belonging only to  $\Gamma_p$  and not to the previously computed ridges. Thus, in that context, we no longer perform the ridge detection until  $P$  ridges are detected, but until a small fraction of the energy is not contained in the vicinity of the detected ridges. In what follows, this RD will be called *redundant ridge detector* (R-RD), to emphasize the fact that some ridges may have some points in common.

An illustration of the practical impact of using R-RD instead of S-RD is given in Figure 3.3 (a) and (b), corresponding to the detection of two ridges with S-RD and to the application of R-RD respectively. As explained above, S-RD enforces the detection of two separate ridges that are not made of LMFs, contrary to R-RD with which the extracted ridge portions can be on top of one another and are necessarily made of LMFs. In the presence of noise, R-RD is slightly adapted so as to ensure the ridges are made of significant LMFs. For that purpose, let  $\tilde{f} = f + \varepsilon$  be a noisy signal, with  $\varepsilon$  a zero mean complex Gaussian white noise. To find out the regions where significant coefficients are located, we consider the hard-thresholded spectrogram  $\mathcal{S}(3)$  that we defined in 2.11. This means that we consider coefficients such that  $|V_{\tilde{f}}^g[n, k]| > 3\hat{\gamma}$  where  $\gamma$  is an estimate of  $\sigma\|h\|_2$ .

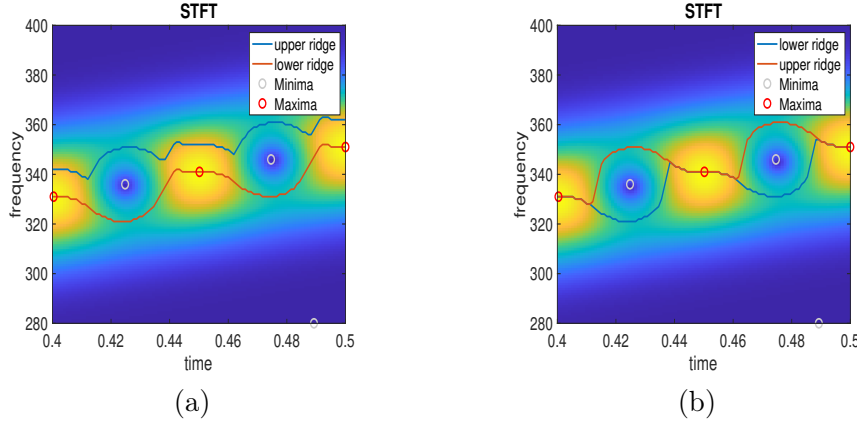


Figure 3.3.: (a): ridges detected using S-RD; (b): ridges detected using R-RD.

In such a noisy context, to build a ridge, one considers the same initialization procedure as previously replacing  $f$  by  $\tilde{f}$  and maximizing only over TF points in  $\mathcal{S}(3)$ , and then the construction follows (3.8) considering only the TF points that are in  $\mathcal{S}(3)$ . All the ridges are extracted iteratively, and the extraction is stopped when the last computed ridge has a length below some threshold. When some noise is present, it is more efficient to put a bound on the ridge length rather than on the remaining energy away from the ridges since, due to noise, this energy can be large. This RD will be called *redundant ridge detector in noise* (R-RDN) in the sequel.

### 3.3.2. Using New RDs to Detect and Localize TFBs

Based on the previous analysis, a simple criterion to investigate whether the proposed new RDs are efficient to detect TFBs is to find out the instances when some ridges have merged. To measure this, let us consider that either with R-RD or R-RDN, depending on whether the signal is noisy or not,  $K$  ridges  $(\Gamma_k)_{k=1, \dots, K}$  have been extracted. Then, let us define the following quantity:

$$R(\sigma) = \begin{cases} 1, & \text{if } \min_{n, 1 \leq i < j \leq K} (|\Gamma_i[n] - \Gamma_j[n]|) = 0 \\ 0, & \text{otherwise.} \end{cases} \quad (3.9)$$

Thus,  $R(\sigma) = 1$  when some ridges have some points in common, and it is equal to zero otherwise. We will see in the Results section that the values of  $R$  computed from the extracted ridges are in accordance with the necessary and sufficient conditions on the existence of TFBs given by the theoretical analysis of Section 3.2.

Once we know two ridges have some points in common in the TF plane, we spot where the merging occurs, the position of which being an estimate of the location of a TFB. An illustration of the procedure is given in Figure 3.4, first row and in the absence of noise (so we use R-RD for ridge detection). The STFT in Figure 3.4 (a) to (c) are computed with  $\sigma = 0.02, 0.03$ , and  $0.04$  respectively. The stars on these figures indicate where some ridges have merged. It is interesting to notice that, as expected, a smaller  $\sigma$  creates more interference between the modes and thus more TFBs. In that case, by increasing  $\sigma$ , it is possible to get three disjoint ridges (see Figure 3.4 (c)), telling us that the separation of the modes will be easier with such a  $\sigma$ . Then, on the second row of Figure 3.4, we give an illustration in the presence of noise, considering the same signal as on the first row, and with  $\sigma = 0.04$  as we know such a value leads to three different ridges in the absence of noise. In that context, we use R-RDN for ridge detection, and consider different noise levels (input SNR of 20, 10 and 5 dB) corresponding to Figure 3.4 (d) to (f). In these figures, we plot TF regions corresponding to  $\mathcal{S}(3)$  and then the ridges computed in these regions using R-RDN. We notice that for the studied noise, the merging of the ridges are located in the same regions as in the noiseless case, when the input SNR is larger than 5 dB, which means the detection of TFBs using the proposed approach seems to be robust to noise. Nevertheless, at this stage, no general conclusions can be drawn since only one realization of the noise is considered, therefore the quality of the detection is going to be investigated more in details in the following section.

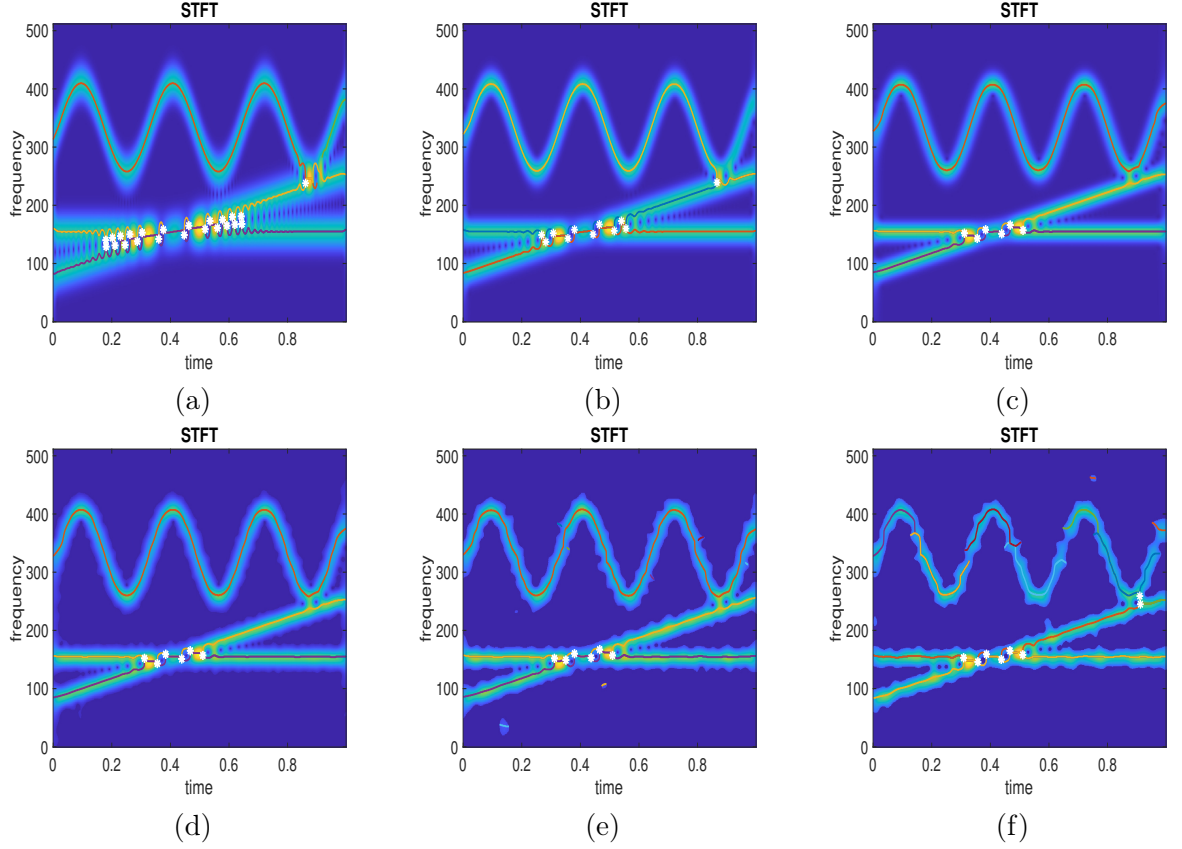


Figure 3.4.: (a): STFT of a three mode signal along with the ridges superimposed, in the definition of the Gaussian window  $\sigma = 0.02$ , the stars corresponds to the locations of the merging of ridges (R-RD is used for ridge detection); (b): same as (a) but with  $\sigma = 0.03$  ; (c): same as (a) but with  $\sigma = 0.04$  ;(d): study of the same signal as in (a) with some white Gaussian added noise (input SNR 20 dB), and  $\sigma = 0.04$  in the definition of the Gaussian window; (e): same as (d) except the input SNR equals 10 dB;(f): same as (d) except the input SNR equals 5 dB.

### 3.4. Results

In this section, we first investigate how the proposed new RDs enable us to detect TFBs. Then, we study whether the localization of the merging of ridges computed with our algorithm is a robust criterion to localize TFBs. Finally we propose an application of the potential use of the localization of TFBs for the analysis of vibrato signals.

### 3.4.1. TFB Detection Based on new RDs

We investigate here how well the merging of ridges with the RDs we propose reflects the presence of TFBs in the TF plane. To do so, we first consider the case of the sum of two pure tones and then the case of two parallel linear chirps. For that purpose we use the signal of Figure 3.1 (a), which corresponds to the sum of two pure tones with the same amplitude. In such a case, we know  $\xi_2 - \xi_1$  and we make  $\sigma$  vary: TFBs are present when  $\alpha = \sqrt{\frac{\pi}{2}}\sigma(\xi_2 - \xi_1) \leq 1$  and not otherwise. This defines a function of  $\sigma$  equal to 1 when a TFB is present and to 0 otherwise. This function is denoted by *ground truth* (GT) hereafter. Then, from (3.9),  $R(\sigma) = 1$  when some ridges have some points in common and is equal to zero otherwise. So  $R$  should be close to GT if it is relevant to detect the presence of TFBs. To check this, In Figure 3.5 (a), we display  $R$  along with GT, in noiseless and noisy cases, R-RD and R-RDN being used respectively to compute the different ridges. We notice that the appearance of TFBs corresponds to  $\sigma$  below a certain threshold, which is computed with a great accuracy using (3.9) for input SNR larger than 20 dB. This threshold is slightly increased when an heavier noise is added to the signal, meaning the noise creates some TFBs that are not present in the noiseless case.

Now, if one considers the signal of Figure 3.1 (c), where the modes are with different amplitude, GT has to be modified because TFBs are present either if  $\alpha \leq 1$  or if  $\alpha > 1$  and  $|\log(A)| \geq -2 \operatorname{arccosh}(\alpha) + 2\alpha\sqrt{\alpha^2 - 1}$ . In Figure 3.5 (b), we display GT, as well as  $R$ , for the just mentioned signal, and we notice that the appearance of TFBs is detected for a very similar value of  $\sigma$  than that given by the theory when little noise is added. As in the previous case, we notice that, for some values of  $\sigma$ , the noise creates some TFBs that are not present in the noiseless case. Note finally that when (3.9) is used to estimate the presence of TFBs when the signal is made of two parallel linear chirps, similar results to those displayed in Figure 3.5 are obtained.

The question we would like to address is whether the TFBs generated by noise is detrimental to the performance of the procedure we propose to localize the TFBs of the noiseless signal.

### 3.4.2. TFBs Localization Based on Ridge Merging

We hypothesized in the previous section that the locations of the merging of ridges were a good indicator of those of TFBs, even in noisy situations. The results depicted on the second row of Figure 3.4 were encouraging in that matter, but should be generalized to many different realizations of the noise to conclude. For that purpose, we again consider

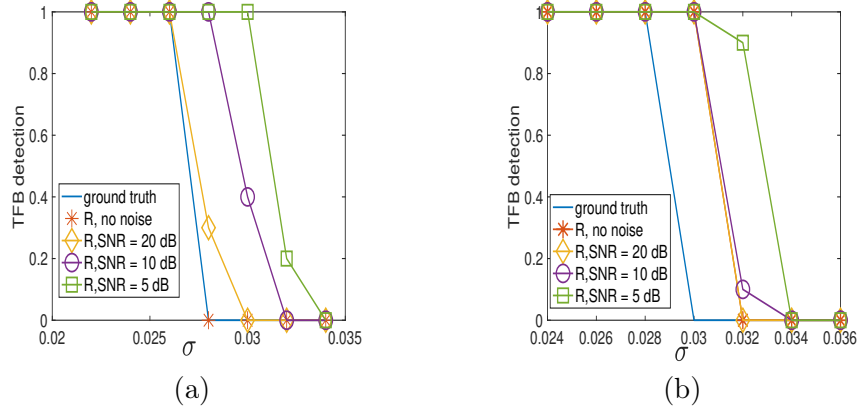


Figure 3.5.: (a): TFB detection for the signal of Figure 3.1 (a) either in the noiseless case or for different noise levels, using the function  $R$ , along with the ground truth; (b): TFB detection for the signal of Figure 3.1 (c) using either  $R$  or the ground truth. The results are averaged over 10 realizations of the noise.

the signal studied in Figure 3.4, and compute the STFT with  $\sigma = 0.04$ , for which, in the absence of noise, the merging of ridges only occurs between the pure tone signal and the linear chirp. To investigate whether this is still the case in the presence of noise, we compute two masks corresponding to where the interference between two modes  $f_1$  and  $f_2$  are significant, namely we define the region of interference of level  $T$  as:

$$\mathcal{I}_T(f_1, f_2) = \left\{ [n, k], \text{ s.t. } |\Re(V_{f_1}^g[n, k]V_{f_2}^g[n, k]^*)| \geq T \right\}. \quad (3.10)$$

We compute these masks when the modes  $f_1$  and  $f_2$  are the pure tone signal and the linear chirp of Figure 3.4, and we refer to that mask as *zone 1*, in what follows. This mask is depicted in Figure 3.6 (a). Then, we carry out the same computation when the modes  $f_1$  and  $f_2$  are the linear chirp and the mode with sinusoidal phase of Figure 3.4, and the obtained mask is referred to as *zone 2*, in what follows (this mask is depicted in Figure 3.6 (b)). Having defined these masks we investigate whether the procedure described in Section 3.3 localizes the merging of ridges mainly in zone 1 when the noise level varies, bearing in mind TFBs are only present in zone 1 in the absence of noise. In Figure 3.6 (c), we display on the top graph the percentage of detections in zone 1, zone 2 and elsewhere in the TF plane (which we denote by false detections). We notice that whatever the noise level, some TFBs are always detected in zone 1, while TFBs are very rarely detected in zone 2 when the noise level is low, and false detections are also very rare. Notice that, though detection in zone 2 and false detections increase when the

noise level increases, these correspond to very few TF points. Indeed, for each realization of the noise we compute the proportion of points corresponding to TFB detections that belong to each region, and remark that only very few TF points are detected outside of zone 1 as illustrated on the bottom graph of Figure 3.6 (c), in which the proportions of TF points in zone 1, zone 2 and elsewhere, are depicted. This means that the proposed procedure is very efficient to localize the TFBs of the original signal.

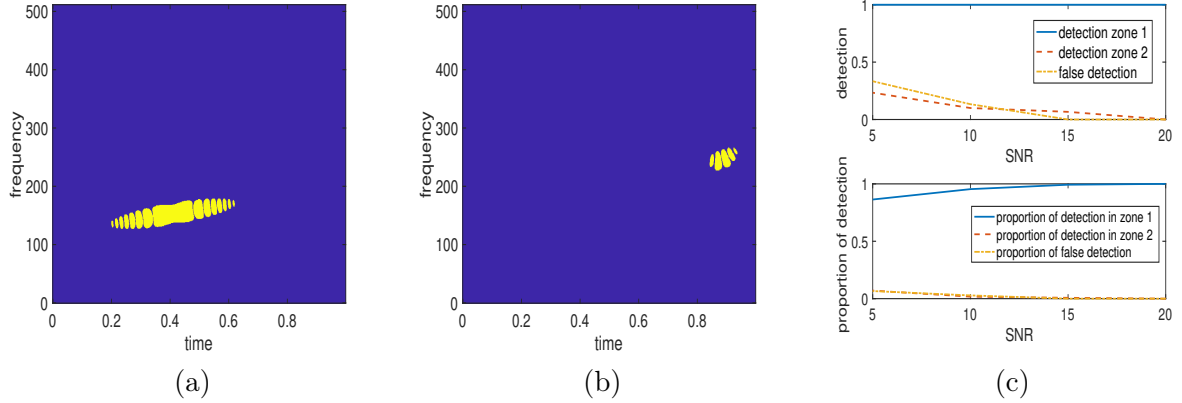


Figure 3.6.: (a): Interference mask associated with the constant and linear chirps of Figure 3.4, denoted by zone 1; (b): Interference mask associated with the linear chirp and the mode with sinusoidal phase of Figure 3.4, denoted by zone 2; (c): top graph: percentage of detection of TFBs in zone 1, zone 2, and elsewhere (false detections) when the noise level varies, bottom graph: percentage of points associated with the detection of TFBs in the different zones when the noise level varies (the number of noise realization equals 30).

### 3.4.3. Application to Voice Signals

In this section, our goal is to give an illustration of the potential interest of the procedure we propose to localize TFBs in practical situations, and in particular for the study of voice signals. Indeed, for such signals, it is of interest to have access to its different harmonics, which can be done by extracting the ridges associated with them in the TF plane. A natural question one then asks is whether this ridge extraction can be trusted. We are going to show that the procedure we propose to localize TFBs offers, in some way, an answer to that question.

To illustrate this, we consider the voice signal of a 22-year-old female taken from the recording 62 of the Saarbrücken Voice Database [65]. This signal contains 126902 samples for a length of 2.538 seconds and for efficiency purpose, we resample it by a factor of 10.



It contains sustained /a/ vowels of the type “low-high-low” meaning there is a change on the pitch of the signal. The STFT of such a signal is depicted in Figure 3.7 (a), (b) and (c) for the same value of  $\sigma$ . In Figure 3.7 (a), we consider that 95 % of the total energy is contained in the vicinity of the ridges when applying R-RD, and in Figure 3.7 (b) 98 % of the energy is contained in the vicinity of the ridges. We notice in Figure 3.7 (a) that the ridge associated with higher frequencies mixes the information corresponding to different harmonics. By considering a higher level of energy contained in the vicinity of the ridges as in Figure 3.7 (b), one detects a TFB at the time instant where the ridge detection was not correct in Figure 3.7 (a). This means that, at this time instant, the frequency resolution is not sufficient to well separate some of the ridges and thus the associated harmonics: the algorithm we propose to localize TFBs tells us where to pay particular attention.

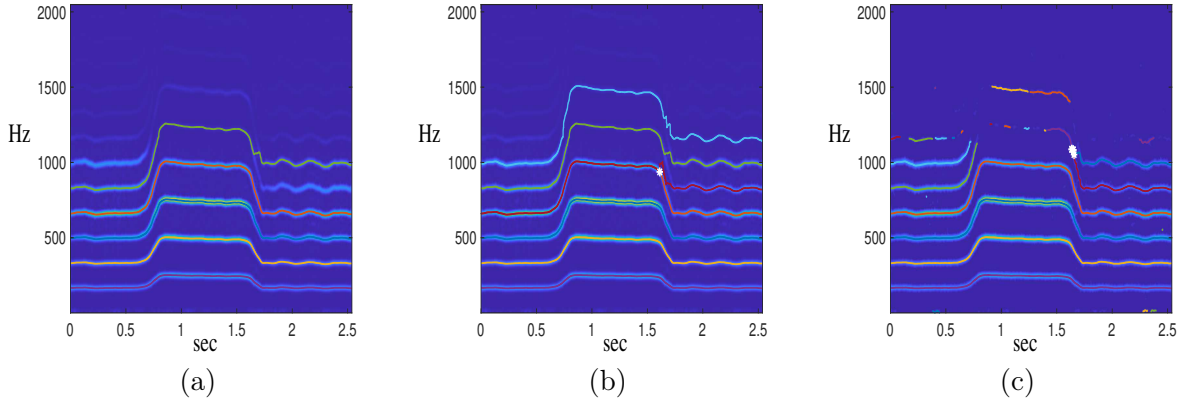


Figure 3.7.: (a): STFT of a voice signal (recording 62 of the Saarbrücken Voice Database [65]) along with the ridges associated with 95% of the total energy (no TFB is detected); (b): same as (a) but when 98% of the total energy is located in the vicinity of the ridges (a TFB corresponding to the white star symbol is detected); (c): same as (a) but in the presence of noise (a TFB is detected in the same TF region as in (b)).

Now in Figure 3.7 (c), we consider a noisy version of the same signal (input SNR 20 dB), compute the spectrogram with the same window as previously, then the ridges using R-RDN and finally localize TFBs. We remark that, due to noise, the ridge associated with higher frequencies in Figure 3.7. (a) is split into several ridge portions but also that a TFB is still detected at the time where interference occurred in Figure 3.7 (b). So, the procedure we propose to localize TFBs is interesting in that it tells us, even in a noisy context, where the frequency (or time) resolution in the spectrogram is not fine enough

to well separate the harmonics.

### 3.5. Analysis of Interference Based on the Zeros of the Spectrogram

In this section, we consider an approach based on the zeros of the spectrogram to localize interference. In some sense, this approach can be seen as dual to the one proposed just above, in that ridges are often associated with high energy TF regions.

The zeros have already been used in TF analysis, for instance, in the context of the auditory system [66], where, using reassignment, the repulsive effect of the zeros is illustrated, or in the context of [67] where the relation between the poles of the phase with the location of the zeros is investigated.

Then in [68], it is explained that the STFT can be viewed as a function of the complex plane, using the Bargmann factorization of the STFT [69, Section 3.4]. It can in fact be regarded as an entire function that is completely characterized by its zeros, which, by construction [68], are the zeros of the STFT.

The Delaunay triangulation based on the spectrogram's zeros is a TFR that can be used for denoising and MR purposes [68]. Succinctly explained, since the distance between two neighboring zeros should not exceed a certain value [68], it is proposed in that paper to consider the Delaunay triangulation of the zeros and then to discriminate the zeros related to the signal from those related to noise by considering a threshold on the length of the triangle edges. To build the signal domain, the following simple procedure is considered [68]: The shape of triangles associated with noise have a specific distribution. When a triangle is an outlier for this distribution, it is associated with signal. Adjacent outliers are then considered belonging to the same mode, enabling MR.

Further mathematical description is given in [70], on statistical properties, notably the link between zeros and Gaussian analytic functions. When one considers a complex and i.i.d. white Gaussian noise  $\varepsilon$ , the zeros of its spectrogram are the set of TF coefficients satisfying  $|V_\varepsilon^g(t, \eta)|^2 = 0$ , and their distribution is uniform on the TF plane [70].

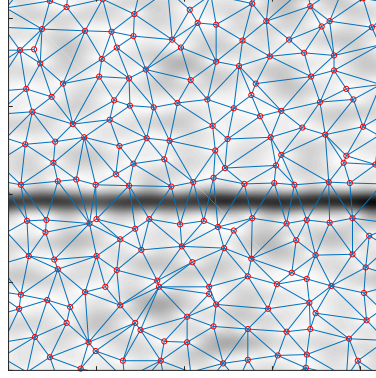


Figure 3.8.: Illustration of the zero based Delaunay triangulation, where the red circles represent the zeros of the spectrogram.

### 3.6. Perspectives for a Zero Based Technique to Identify Interference

In this section, we study how to identify the signal domain in the TF plane, using the zeros of the spectrogram. As explained in the previous section, this has been done in [68], but in our case, we also want to identify interference in the TF plane, meaning we want to split the TF plane into three categories associated with either noise, interference or signal. We also consider a different approach, where we study the Voronoi cells associated with the zeros instead of their Delaunay triangles. To illustrate the perspectives described in this section, we consider the signal  $\tilde{f} = f_1 + f_2 + \varepsilon$ , where  $f_1$  and  $f_2$  are two *pure harmonics* (PHs) that interfere with each other, and  $\varepsilon$  is a complex and i.i.d. white Gaussian noise. In Figure 3.9 (a), we depict the Voronoi cells of the zeros of the modulus of the STFT, namely  $|V_{\tilde{f}}^g|$ , where  $g$  is a Gaussian window.

In order to separate the TF plane, we aim at identifying three types of zeros associated with one category. We propose to discriminate them using the distribution of energy in their Voronoi cells; such a cell is denoted by  $C_{\tilde{f}}^i$ , and is a polygon composed of  $J_i$  segments  $C_{\tilde{f}}^{i,j}$  for  $j \in \{1, \dots, J_i\}$ . Figures 3.9 (b), (c) and (d) respectively illustrate Voronoi cells associated with noise, interference and the signal. In these figures, one can observe that the energy in a segment can be significantly larger when it crosses signal information. Taking into account these observations, our goal is to study the maximum energy on a segment  $C_{\tilde{f}}^{i,j}$ , defined by

$$M_{\tilde{f}}^{i,j} = \max_{(t,\eta) \in C_{\tilde{f}}^{i,j}} |V_{\tilde{f}}^g(t, \eta)|^2. \quad (3.11)$$

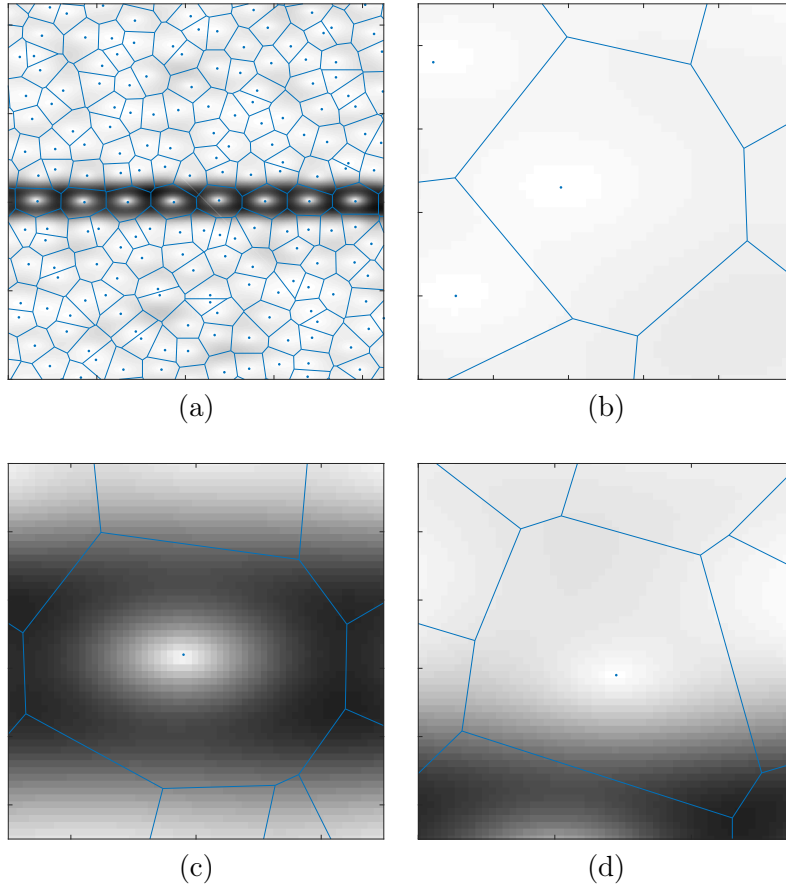


Figure 3.9.: (a): Magnitude of the STFT of two PHs affected by noise, with superimposed Voronoi cells; (b): Zoom on (a) on a Voronoi cell associated with noise; (c): Zoom on (a) on a Voronoi cell associated with interference; (d): Zoom on (a) on a Voronoi cell associated with signal.

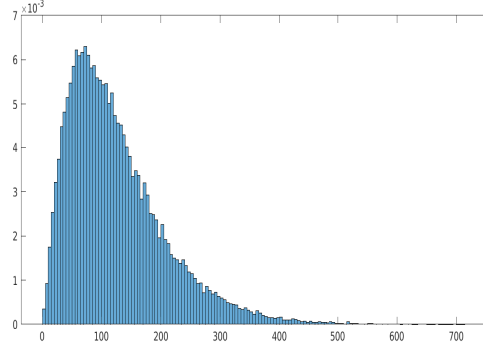


Figure 3.10.: Histogram associated with the selection of the maximum value on  $\approx 80\,000$  segments of Voronoi cells in the case of pure noise.

In order to identify the type of cell  $C_f^i$ , we assume that there exist a threshold  $\mathcal{E}$  such that if  $M_f^{i,j} > \mathcal{E}$ , then  $C_f^{i,j}$  crosses the signal domain. Then, by defining

$$E_i^{max} = \max_{j \in \{1, \dots, J_i\}} M_f^{i,j} \quad (3.12)$$

$$E_i^{min} = \min_{j \in \{1, \dots, J_i\}} M_f^{i,j}, \quad (3.13)$$

we propose to separate the cells by utilizing the following scheme.

If  $E_i^{min} > \mathcal{E}$ ,  $C_f^i$  is related to interference, if  $E_i^{max} \leq \mathcal{E}$ ,  $C_f^i$  is related to noise and otherwise,  $C_f^i$  is related to signal. However, the choice of  $\mathcal{E}$  is delicate, and to set it in a meaningful way, we propose to study the distribution  $M_\varepsilon^{i,j}$ , meaning we only consider pure noise.

A similar problem was dealt with in [71], where the distribution of the spectrogram maxima was identified to follow a Gumbel law. In our case, we restrict our selection of the maximum in a segment of a cell. Namely, first look to determine the law associated with  $M_f^{i,j}$ . Using numerical computation, we depict the histogram of its probability density function in Figure 3.10 and plan identify the law and parameters in a near future.

### 3.7. Conclusion

In this chapter, we studied two different approaches on interference. The first is based on TFBs, for which we carried out the study of the mathematical conditions for their appearance. Then, we introduced a technique to detect TFBs for a general MCS and

use numerical validation to show its accuracy in noisy conditions. We also performed its validation on a real signal, which illustrates that the technique can be used as an indicator of the ridge detection quality.

In the last part of the chapter, we study a different approach based on the zeros of the spectrogram. We proposed a scheme that uses the distribution of the energy around the Voronoi cells associated with the zeros to determine TF regions associated with noise, interference or signal. Assuming that the maxima of the energy in the border of a Voronoi cells and the maxima of the spectrogram are similarly distributed, a new technique could be designed to separate the TF regions, which we plan to investigate in the future.



## 4. Towards Adaptive and Robust Synchrosqueezing

### 4.1. Introduction

In this chapter, we first recall that to get closer to the *ideal time frequency* representation when the modes of a MCS have strong frequency modulation, FSST2 was generalized to any order  $N$  through FSSTN [61]. While to use a  $N$  larger than 2 can be profitable in some instances, it is hard to determine which  $N$  is the best in a particular situation. Ideally, one would seek to find  $N$  in accordance with the order of the phase [72], but the latter is unknown in general. Furthermore, FSSTN, with  $N$  large, is sensitive to noise [72]. In this chapter, we design a technique that aims to determine the best  $N$  locally in the TF plane. Our approach is based on energy concentration, which we interpret as the quality of the reassignment performed by FSSTN. Note that in [73], a multitapered wavelet based SST was introduced to obtain better concentrated TFR and represents an alternative to our technique.

### 4.2. Definition of FSSTN

The objective of this section is to recall the definition of FSSTN, introduced in [20, 61]. To motivate this work, we mention that even though FSST2 leads to a concentrated TFR for linear chirps with Gaussian modulated amplitudes, it is not perfect for a general MCS. For instance, if a mode has a polynomial phase of order greater than two, then one needs a transform of higher order to obtain an accurate reassigned transform, this is precisely the aim of FSSTN.

The definition of FSSTN is based on a Taylor expansion of the signal's amplitude and phase. Let us consider the single mode signal  $f(t) := A(t)e^{2i\pi\phi(t)}$ , where we assume

---

This chapter is based on the work published in [42] (to appear).



$\log(A)$  and  $\phi$  to be polynomial of degree  $N$ . Let  $t \in \mathbb{R}$ , the  $N$ -th order expansion of  $A$  and  $\phi$  at  $\tau$  is

$$f(\tau) = \exp \left( \sum_{k=0}^N \frac{(\tau - t)^k}{k!} \left\{ [\log(A(t))]^{(k)} + 2i\pi\phi^{(k)}(t) \right\} \right). \quad (4.1)$$

Considering  $g \in L^2(\mathbb{R})$ , if  $f \in L^2(\mathbb{R})$ , its STFT satisfies,

$$V_f^g(t, \eta) = \int_{\mathbb{R}} \exp \left( \sum_{k=0}^N \frac{\tau^k}{k!} \left\{ [\log(A(t))]^{(k)} + 2i\pi\phi^{(k)}(t) \right\} \right) g(\tau) e^{-2i\pi\tau\eta} d\tau \quad (4.2)$$

By differentiating with respect to  $t$  and then by dividing by  $2\pi V_f^g$  one obtains,

$$\tilde{\omega}_f(t, \eta) := \frac{\partial_t V_f^g(t, \eta)}{2i\pi V_f^g(t, \eta)} = \sum_{k=1}^N r_k^{[N]}(t) \frac{V_f^{t^{k-1}g}(t, \eta)}{V_f^g(t, \eta)}, \quad (4.3)$$

where

$$r_k^{[N]}(t) = \frac{1}{(k-1)!} \left( \frac{[\log(A(t))]^{(k)}}{2i\pi} + \phi^{(k)}(t) \right). \quad (4.4)$$

In order to define the  $N$ -th order local IF estimate, we first set,

$$\tilde{\omega}_f^{[N]}(t, \eta) = \tilde{\omega}_f(t, \eta) - \sum_{k=2}^N r_k^{[N]}(t) \frac{V_f^{t^{k-1}g}(t, \eta)}{V_f^g(t, \eta)} = r_1^{[N]}(t) \quad (4.5)$$

and then define  $\hat{\omega}_f^{[N]}(t, \eta) := \Re\{\tilde{\omega}_f^{[N]}(t, \eta)\}$  where in our case,  $\hat{\omega}_f^{[N]}(t, \eta) = \phi'(t)$ . A simple way to compute  $r_1^{[N]}$  is first, using (4.3), to remark that

$$\frac{1}{2i\pi} \partial_t V_f^g(t, \eta) = \sum_{k=1}^N r_k^{[N]}(t) V_f^{t^{k-1}g}(t, \eta), \quad (4.6)$$

and that  $\partial_\eta V_f^g(t, \eta) = -2i\pi V_f^{tg}(t, \eta)$ . Then, we deduce the following equality [77],

$$\frac{1}{2i\pi} \begin{bmatrix} \partial_t V_f^g \\ \frac{i}{2\pi} \partial_\eta \partial_t V_f^g \\ \vdots \\ \frac{i^{N-1}}{(2\pi)^{N-1}} \partial_\eta^{N-1} \partial_t V_f^g \end{bmatrix} = \begin{bmatrix} V_f^g & V_f^{tg} & \cdots & V_f^{t^{N-1}g} \\ V_f^{tg} & V_f^{t^2g} & \cdots & V_f^{t^Ng} \\ \vdots & \vdots & \ddots & \vdots \\ V_f^{t^{N-1}g} & V_f^{t^Ng} & \cdots & V_f^{t^{2(N-1)}g} \end{bmatrix} \begin{bmatrix} r_1^{[N]} \\ r_2^{[N]} \\ \vdots \\ r_N^{[N]} \end{bmatrix} = DR. \quad (4.7)$$

Finally, based on simple properties of the determinant of matrices, setting,

$$M_1 = \begin{bmatrix} \partial_t V_f^g & V_f^{tg} & \dots & V_f^{t^{N-1}g} \\ \frac{i}{2\pi} \partial_\eta \partial_t V_f^g & V_f^{t^2g} & \dots & V_f^{t^Ng} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{i^{N-1}}{(2\pi)^{N-1}} \partial_\eta^{N-1} \partial_t V_f^g & V_f^{t^Ng} & \dots & V_f^{t^{2(N-1)}g} \end{bmatrix}, \quad (4.8)$$

one obtains

$$r_1^{[N]} = \frac{1}{2i\pi} \det(D^{-1}) \det(M_1) = \frac{1}{2i\pi} \frac{\det(M_1)}{\det(D)}. \quad (4.9)$$

Then, as  $\partial_t V_f^g = i2\pi\eta V_f^g - V_f^{g'}$ , one gets, for any  $l \geq 1$ :

$$\partial_\eta^l \partial_t V_f^g = (-2i\pi)^l \left( -k V_f^{t^{l-1}g} - V_f^{t^l g'} + 2i\pi\eta V_f^{t^l g} \right), \quad (4.10)$$

leading to  $\det(M_1) = i2\pi\eta \det(D) - \det(U_1) - \det(V_1)$  with

$$U_1 = \begin{bmatrix} 0 & V_f^{tg} & \dots & V_f^{t^{N-1}g} \\ V_f^g & V_f^{t^2g} & \dots & V_f^{t^Ng} \\ \vdots & \vdots & \ddots & \vdots \\ (N-1)V_f^{t^{N-2}g} & V_f^{t^Ng} & \dots & V_f^{t^{2(N-1)}g} \end{bmatrix}, \quad (4.11)$$

$$V_1 = \begin{bmatrix} V_f^{g'} & V_f^{tg} & \dots & V_f^{t^{N-1}g} \\ V_f^{tg'} & V_f^{t^2g} & \dots & V_f^{t^Ng} \\ \vdots & \vdots & \ddots & \vdots \\ V_f^{t^{N-1}g'} & V_f^{t^Ng} & \dots & V_f^{t^{2(N-1)}g} \end{bmatrix}, \quad (4.12)$$

and thus

$$\widehat{\omega}_f^{[N]} = \Re \left\{ r_1^{[N]} \right\} = \eta - \frac{1}{2\pi} \Im \left\{ \frac{\det(U_1) + \det(V_1)}{\det(D)} \right\}. \quad (4.13)$$

Finally, FSSTN is defined by [20, 61],

$$T_f^{g,N}(t, \eta) := \int_{\mathbb{R}^+} V_f^g(t, \tau) \delta \left( \eta - \widehat{\omega}_f^{[N]}(t, \tau) \right) d\tau. \quad (4.14)$$

And, similarly to (1.16, page 4), FSSTN is invertible,

$$f(t) = \frac{1}{g(0)} \int_{\mathbb{R}} T_f^{g,N}(t, \eta) d\eta. \quad (4.15)$$

### 4.3. Adaptive selection of the order

The objective of this section is to set the order  $N$  of FSSTN locally, to get a sharp TFR. In the general case, one does not know the signal's modulation, and therefore it is not possible to set the order of FSSTN to the degree of the phase. An adaptive way to set  $N$  would be to measure the concentration of energy, and find  $N$  that maximizes it.

Let us first consider the general case of a noisy MCS affected by the same kind of white Gaussian noise as in Chapter 2.

$$\tilde{f} = f + \varepsilon \quad (4.16)$$

where  $f$  is a MCS in  $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  of  $P$  modes, and  $\varepsilon$  a white Gaussian noise. Our interest is to concentrate the energy to ease the reading of the signal in the TF plane, and for that purpose, we consider that for each mode  $p$ , its ridge  $\Gamma_p$  is given by a RD. We then measure the concentration of energy on the ridge in  $T_{\tilde{f}}^{g,N}$ . In the ideal case, considering a monocomponent signal  $f(t)$ , based on the reconstruction formula of FSSTN, all the information is located on the ridge,

$$f(t) \approx \frac{T_{\tilde{f}}^{g,N}(t, \Gamma(t))}{g(0)}. \quad (4.17)$$

Based on that, we measure the energy concentration using the following estimate of the instantaneous amplitude, when  $g(0) > 0$ ,

$$E_N^p(t) = \frac{|T_{\tilde{f}}^{g,N}(t, \Gamma_p(t))|}{g(0)}, \quad (4.18)$$

where  $g$  is a window in  $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  and RD is performed on  $T_{\tilde{f}}^{g,N}$ . With a continuous ridge, where in the ideal case  $\Gamma_p(t) = \phi'_p(t)$ , the selection of  $N$  could be made by maximizing (4.18), by introducing

$$\mathcal{N}^p(t) = \arg \max_{N \in \mathbb{N}} E_N^p(t). \quad (4.19)$$

For the sake of evaluating the relevance of this method, using the invertibility of FSSTN,

we define,

$$R_{f_p}^{\mathcal{N}^p}(t) := \frac{T_{\tilde{f}}^{g, \mathcal{N}^p(t)}(t, \Gamma_p(t))}{g(0)}, \quad (4.20)$$

which is an estimate of the  $p$ -th mode, namely  $f_p(t) \approx R_{f_p}^{\mathcal{N}^p}(t)$ . This enables us to make the comparison of this technique with the signal in terms of SNR. Before that, in order to proceed with the practical study of this method, we first detail the corresponding discrete framework.

We consider the sequence  $\tilde{f}[n] = \tilde{f}(\frac{n}{L})$ , where  $n \in \{1, \dots, L\}$ . The discrete FSSTN definition is given by,

$$T_{\tilde{f}}^{g, N}[n, w] = \sum_{k=1}^J V_{\tilde{f}}^g[n, k] \delta\{w, \lfloor \hat{\omega}_{\tilde{f}}^{[N]}[n, k] \rfloor\} \quad (4.21)$$

where  $\lfloor x \rfloor$  is the closest integer to  $x \in \mathbb{R}$  and  $\delta$  is the Kronecker symbol. We also recall that the discrete version of  $\hat{\omega}_{\tilde{f}}^{[N]}$  used in 4.21 only uses the discrete formalism of the STFT.

However, by using this transform directly, instabilities related to the noise and numerical computations may significantly worsen the quality of mode reconstruction with our method. Indeed, when  $\hat{\omega}_{\tilde{f}}^{[N]}[n, k]$  is located between two frequency bins, it is likely that the energy will be dispatched in the upper and lower bins. Obviously, this will impact the measure given by (4.18). To avoid this effect, we consider a variant of  $T_{\tilde{f}}^{g, N}$ , where we shift the frequency grid by  $q \in \{0, \frac{1}{2}\}$  of a frequency bin and choose  $q$  that maximizes the concentration. For that purpose, we define the shifted variant of FSSTN by,

$$T_{\tilde{f}, q}^{g, N}[n, w] := \sum_{k=1}^J V_{\tilde{f}}^g[n, k] \delta\{w, \lfloor \hat{\omega}_{\tilde{f}}^{[N]}[n, k] + q \rfloor\}, \quad (4.22)$$

where  $J$  is the number of frequency bins. Then a second issue is that we cannot compute FSSTN for all  $N \in \mathbb{N}$ , and therefore we set  $N_0 \in \mathbb{N}$  and consider orders up to  $N_0$ . Taking the shift into account, we define the energy concentration for  $N \in \{1, \dots, N_0\}$  in the discrete framework by,

$$E_N^p[n] := \max_{q \in \{0, \frac{1}{2}\}} \frac{|T_{\tilde{f}, q}^{g, N}[n, \Gamma_p[n]]|}{g(0)}, \quad (4.23)$$

which represents an estimate of the instantaneous amplitude in the discrete case. Note that the ridge given by  $\Gamma_p$  is also computed on  $T_{\tilde{f}, q}^{g, N}$ , and that as a consequence, the shift is fixed independently for each mode.

In that context, we set  $\mathcal{N}^p[n] = \arg \max_{N \in \{1, \dots, N_0\}} E_N^p[n]$ . This allows us to set the proper choice of  $q$  depending on  $n$  using,

$$Q_p[n] := \arg \max_{q \in \{0, \frac{1}{2}\}} \left| T_{\tilde{f}, q}^{g, N}[n, \Gamma_p[n]] \right|. \quad (4.24)$$

Finally, to approximate the mode, as in (4.20), we set,

$$R_{f_p}^{\mathcal{N}^p}[n] = \frac{1}{g(0)} T_{\tilde{f}, Q_p[n]}^{g, \mathcal{N}^p(t)}[n, \Gamma_p[n]]. \quad (4.25)$$

To draw conclusions on the quality of our method, we can now compute the output SNR using this estimate of  $f_p$ .

#### 4.3.1. Numerical Results for the Monocomponent Case

In this subsection, we consider monocomponent signals with constant amplitude  $A = 1$ . We first look at amplitude estimates given by  $E_N$ , for  $N \in \{2, \dots, 4\}$ . Note that first order is not considered, as it does not take any modulation into account, and is significantly worse than other orders.

In Figure 4.1 (a), the signal has a high-order and oscillating frequency. From 4.1 (b), we remark that  $E_4$  is stable around one and is the best estimate for this signal. In Fig

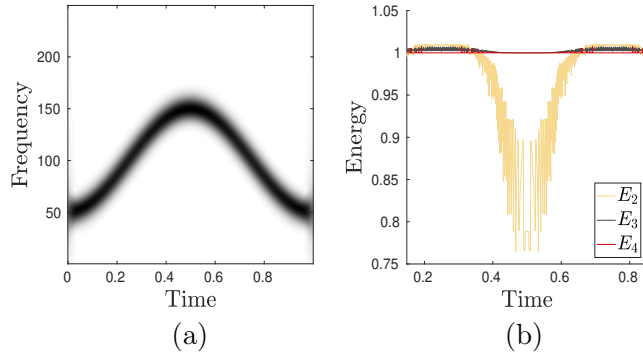


Figure 4.1.: (a): magnitude of the STFT of a mode with an oscillating phase; (b):  $E_N$  for  $2 \leq N \leq 4$  for the signal whose STFT is depicted in (a)

4.2 (a), we consider a signal that is more modulated than the previous one. In this case,  $E_2$ ,  $E_3$  and  $E_4$ , are respectively depicted on Figure 4.2 (b), (c) and (d). This time, the best choice is not as obvious because  $E_4$  creates overshoots, but still, on average, the estimation is much better. Such overshoots are explained by time interference that locally increase the energy.

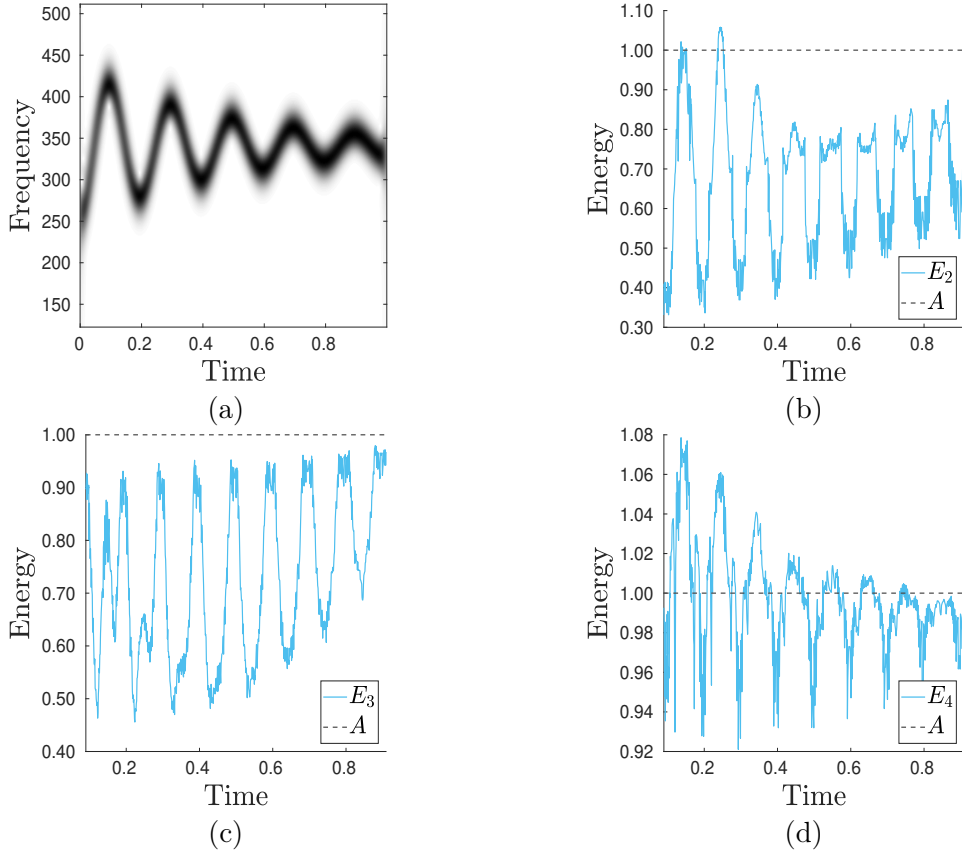


Figure 4.2.: (a): STFT magnitude of a mode with damped sinusoidal frequency and amplitude  $A = 1$ ; (b),(c),(d):  $E_2, E_3$  and  $E_4$  for the mode corresponding to (a), respectively.

In Figure 4.3, we proceed with the comparison between the adaptive selection of the order for  $R_f^N$  and  $R_f^N$  measured in terms of output SNR associated with mode reconstruction when the level of noise varies. It is clear that the adaptive technique allows good reconstruction quality. We also note that when the noise level is strong,  $R_f^N$  gives similar results to FSST2.

We next focus on the results for MCSs.

#### 4.3.2. Numerical Results for the Multicomponent Case

In this subsection, we illustrate the benefits of adaptive selection of the orders for each mode.

For that purpose, we consider the signal of Figure 4.4 (a), which contains two modes where, this time, the amplitude varies. The output SNR for the different techniques is depicted in Figure 4.4 (b). As expected, using  $R_{f_p}^{N_p}$  leads to a higher output SNR, which

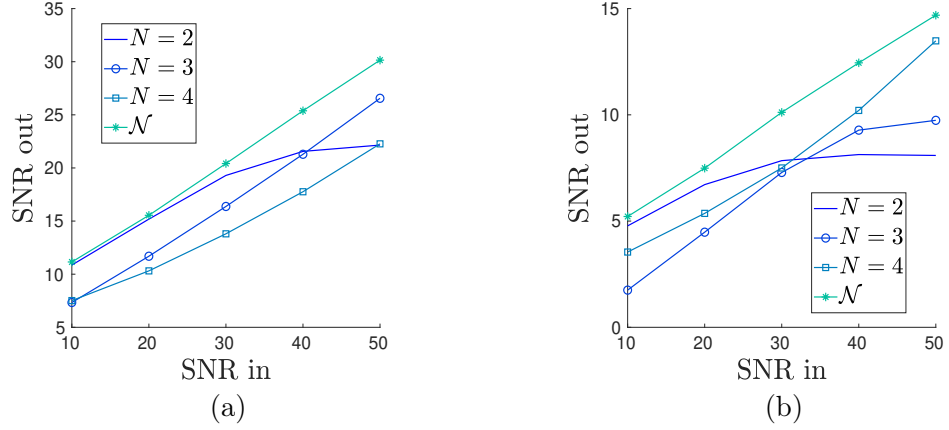


Figure 4.3.: (a): output SNR corresponding to  $\text{SNR}(f, R_f^N - f)$  for  $N = 2, \dots, 4$  or to  $\text{SNR}(f, R_f^N - f)$  for the signal of Figure 4.1 (a); (b): same as (a) but for the signal of Figure 4.2 (a).

means that an adaptive selection of the order depending on time is beneficial compared with using FSSTN with a fixed order  $N$ . The rationale for such a behavior is that with, in the case of MCSs, the choice of the order can be set independently for each mode. Finally, we also note that when the input SNR is around 10 dB or below, fixing  $N = 2$ , namely considering  $R_{f_p}^2$  is competitive, like in the monocomponent case presented in previous the subsection.

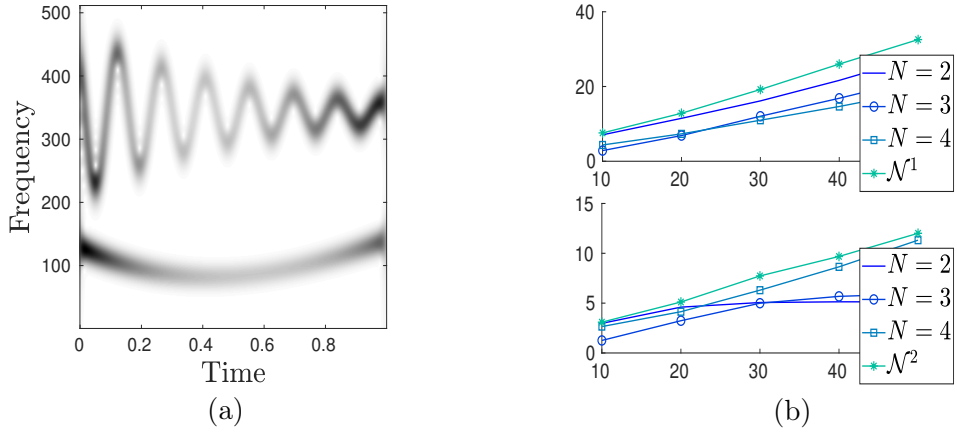


Figure 4.4.: (a): modulus of the STFT of the sum of two modes with varying amplitude; (b): mode reconstruction results as a function of input SNR.

### 4.3.3. Application to a Bat Echolocation

We now investigate how the proposed adaptive technique improves the quality of reconstruction on a bat echolocation call whose STFT magnitude is depicted in Figure 4.5 (a). In our study, we only look for the three most energetic modes in ridge detection and put  $R_f^{\mathcal{N}_a} := \sum_{p=1}^3 R_{f_p}^{\mathcal{N}_p}$ ,  $\mathcal{N}_a$  meaning we consider an adaptive  $N$  in the MCS case. Similarly to  $R_f^{\mathcal{N}_a}$ , we respectively denote signal reconstruction from the three detected modes for,  $1 \leq N \leq 4$ , by  $R_f^N$ . We then compare  $R_f^{\mathcal{N}_a}$  and  $R_f^N$ , for different  $N$ , to  $f$  in terms of SNR, and we again notice that the first technique is much better than the others (see the Table of Figure 4.5 (b)).

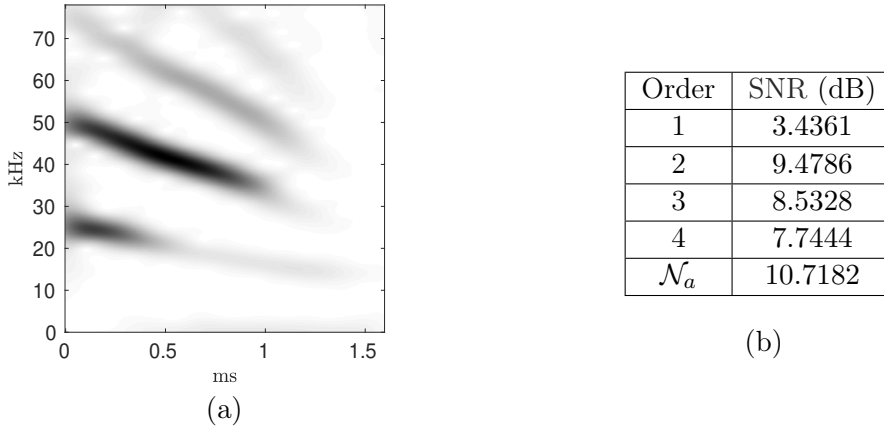


Figure 4.5.: (a): STFT modulus of a bat echolocation call; (b): SNR corresponding to signal reconstruction from the coefficients on the ridges of the 3 most energetic detected modes.

## 4.4. Conclusion

In this chapter, we defined an indicator of energy concentration that we used in order to select the appropriate order of FSSTN to improve TFR readability. To assess this improvement, we measured the quality of mode reconstruction considering a fixed or adaptive order in synchrosqueezing transforms. Numerical results showed that the adaptive technique is of better quality in general, but that FSST2 is also a good choice in very noisy conditions.





## 5. Study of the Chirp Rate Estimate in Case of Noise

In this chapter, the objective is to estimate the CR, in a way that is robust to noise. Such an estimate is of interest, notably in synthetic aperture radar and inverse synthetic aperture radar imaging [74], and in voice signals to study the jitter phenomenon [75, 76]. We focus on the CR estimator in FSST2 and first study its behavior in noisy conditions. In the previous chapter, the numerical simulations showed that the second order is relevant to consider when the noise is strong, and enables good energy concentration, and that to consider a higher order for FSSTN does not improve the result when the input SNR is around 10 dB or below. When one performs a study in very noisy conditions, FSST2 is a valid choice of representation. However, in Chapter 2, we highlighted that this estimate is sensitive to noise and required further investigations. With that in mind, we study how the noise damages the CR estimate and show that the estimation error is oscillatory. Then, we detail a technique to filter out most of these oscillations using the power spectral density of the error. We finally evaluate the quality of the filtered estimate and its stability to noise on a real voice signal.

To carry out this study, a general definition of the STFT is required, because the noise is not a Lebesgue-integrable function. With a proper definition, the STFT of the noise can be considered as a random variable. For that purpose, we will consider that  $\varepsilon$  is a tempered distribution, namely  $\varepsilon \in S'(\mathbb{R})$ , and define the STFT by [70],

$$V_\varepsilon^g(t, \eta) = \int_{\mathbb{R}} g(\tau - t) e^{-2i\pi(\tau - t)\eta} dB_\varepsilon(\tau), \quad (5.1)$$

where  $g \in S(\mathbb{R})$ . The right-hand side is an Itô integral with respect to the one dimensional Brownian motion  $B_\varepsilon(\tau)$ . The derivative of Brownian motion is a white Gaussian noise and therefore one can think of  $dB_\varepsilon(\tau)$  as  $\varepsilon(\tau)d\tau$ . It follows that  $V_\varepsilon^g(t, \eta)$  has a covariance given by the 2-by-2 square matrix  $\sigma_\varepsilon^2 \|g\|_2^2 I_2$  [78].

---

This chapter is based on the work published in [43].

We now proceed with the analysis of the CR estimate for a monocomponent signal  $f$ , setting  $\tilde{f} = f + \varepsilon$  and  $g(t) = e^{-\pi \frac{t^2}{\sigma^2}}$ . Remarking that  $g \in S(\mathbb{R})$  and using the properties of  $g$ , we compute a simpler expression of  $\hat{q}_{\tilde{f}}$ . For the sake of clarity, we omit time-frequency coordinates  $(t, \eta)$  in the most complicated expressions, and first remark that in our context, one has

$$\hat{q}_{\tilde{f}} = -\frac{1}{2\pi} \Im \left\{ \frac{(V_{\tilde{f}}^g)^2}{V_{\tilde{f}}^g V_{\tilde{f}}^{t^2 g} - (V_{\tilde{f}}^{tg})^2} \right\}. \quad (5.2)$$

Indeed, in equation (1.17, page 4) with our choice for  $g$ , the STFTs depending on  $g'$  can be simplified using  $V_{\tilde{f}}^{g'}(t, \eta) = -2\frac{\pi}{\sigma^2} V_{\tilde{f}}^{tg}(t, \eta)$ . Having a simple expression of  $\hat{q}_{\tilde{f}}$ , our interest is now to denoise it.

### 5.0.1. Study of a Simplified CR Estimate

Now, we further simplify  $\hat{q}_{\tilde{f}}$  by identifying negligible terms and thus defining new estimators. In Section 2.5, we detailed that the STFT of a linear chirp is a Gaussian centered at  $\phi'(t)$  with respect to frequency. This means that its frequency derivative vanishes at  $\phi'(t)$ . Recalling that,

$$\frac{-\partial_{\eta}}{2i\pi} V_f^g(t, \eta) = V_f^{tg}(t, \eta), \quad (5.3)$$

and assuming that  $f$  can be approximated by a linear chirp, we use that  $V_f^{tg}(t, \phi'(t)) \approx 0$ , and simplify the operator by making a Taylor expansion, in the vicinity of the ridge.

$$\hat{q}_f = -\frac{1}{2\pi} \Im \left\{ \frac{V_f^g}{V_f^{t^2 g}} \frac{1}{1 - \frac{(V_f^{tg})^2}{V_f^g V_f^{t^2 g}}} \right\} \approx -\frac{1}{2\pi} \Im \left\{ \frac{V_f^g}{V_f^{t^2 g}} \right\}. \quad (5.4)$$

From (5.4), we define the new estimate based on the last approximation by,

$$\bar{q}_f := -\frac{1}{2\pi} \Im \left\{ \frac{V_f^g}{V_f^{t^2 g}} \right\}. \quad (5.5)$$

Taking the noise into account, one gets,

$$\bar{q}_{f+\varepsilon} = -\frac{1}{2\pi} \Im \left\{ \frac{V_f^g + V_{\varepsilon}^g}{V_f^{t^2 g} + V_{\varepsilon}^{t^2 g}} \right\}. \quad (5.6)$$

Numerically, we show that  $\bar{q}_{f+\varepsilon}$  varies similarly to  $\hat{q}_{f+\varepsilon}$  for the three signals of Figure 5.1 (a), (b) and (c), as illustrated on Figure 5.1 (d), (e) and (f). We also notice that the quality of the estimation changes with the modulation of the signal. In order to denoise  $\bar{q}_{f+\varepsilon}$ , we are going to study it as a random variable. We first study the bias of  $\Re\{\bar{q}_{f+\varepsilon}(t, \eta)\}$  as an estimate of  $\phi''(t)$ .

**Proposition 5.0.1.** *At a STFT ridge point  $(t, \phi'(t))$  of a linear chirp one has:*

$$\mathbb{E}\{\bar{q}_{f+\varepsilon}\} = \mathbb{E}\left\{-\frac{1}{2\pi}\Im\{Z\}\right\} = \phi''(t) \left(1 - e^{-\frac{|V_f^{t^2g}|^2}{\sigma_\varepsilon^2 \|t^2g\|^2}}\right),$$

where  $\sigma_\varepsilon^2$  is the variance of the noise and  $Z(t, \eta) = \frac{V_{f+\varepsilon}^g(t, \eta)}{V_{f+\varepsilon}^{t^2g}(t, \eta)}$ .

The proof is given in Appendix B.1. We remark that the bias depends on  $V_f^{t^2g}$ , which we develop below to study its effect. In the general case, one has [77]

$$\begin{aligned} V_f^{t^k g}(t, \eta) &= \left(\frac{i}{2\pi}\right)^k \partial_\eta^k V_f^g(t, \eta) \\ &= \frac{-\sigma^2(1 + i\phi''(t)\sigma^2)}{1 + (\phi''(t)\sigma^2)^2} \left[ i(\eta - \phi'(t))V_f^{t^{k-1}g}(t, \eta) - \frac{k-1}{2\pi}V_f^{t^{k-2}g}(t, \eta) \right], \end{aligned} \quad (5.7)$$

and thus,

$$V_f^{t^k g}(t, \phi'(t)) = \frac{\sigma^2(1 + i\phi''(t)\sigma^2)}{1 + (\phi''(t)\sigma^2)^2} \frac{k-1}{2\pi} V_f^{t^{k-2}g}(t, \phi'(t)), \quad (5.8)$$

which implies that [19, proposition 1] (omitting  $(t, \phi'(t))$  for the sake of simplicity),

$$|V_f^{t^2g}| = \frac{\sigma^2 |V_f^g|}{2\pi \sqrt{1 + (\phi''(t)\sigma^2)^2}} = \frac{A\sigma^2}{2\pi(1 + (\phi''(t)\sigma^2)^2)^{3/4}}. \quad (5.9)$$

Then since  $\|t^2g\|^2 = \sigma^5 \int \tau^4 e^{-2\pi\tau^2} d\tau = \frac{3\sigma^5}{(2\pi)^{24}\sqrt{2}}$ , we deduce that,

$$\frac{|V_f^{t^2g}|^2}{\|t^2g\|^2} = \frac{A^2 4\sqrt{2}}{3\sigma(1 + (\phi''(t)\sigma^2)^2)^{3/2}}. \quad (5.10)$$

This means that the bias is growing with the modulation. In the case of the signals we depict in Figure 5.1 (a), (b) and (c), the bias is negligible, meaning that  $\mathbb{E}\left\{\frac{|V_\varepsilon^{t^2g}|}{|V_f^{t^2g}|}\right\} = \frac{\sigma_\varepsilon \|t^2g\|}{|V_f^{t^2g}|}$  is small. We use this opportunity to simplify the estimate even more, using the

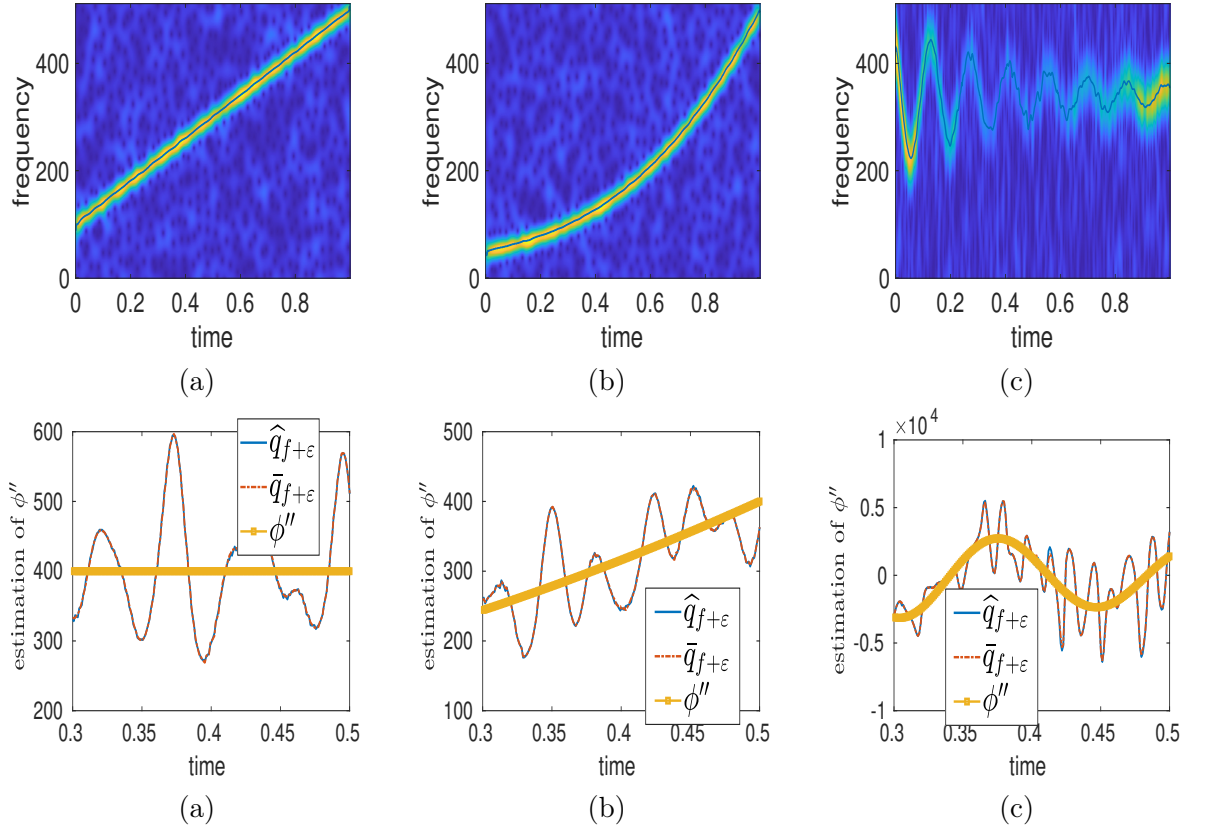


Figure 5.1.: (a): STFT modulus of a noisy linear chirp with the STFT ridge superimposed; (b): same as (a) but for a noisy chirp with fourth order polynomial phase; (c): same as (a) but for a noisy mode with oscillating phase; (d): CR estimator  $\hat{q}_{f+\epsilon}$  or  $\bar{q}_{f+\epsilon}$  estimated on the STFT ridge of the noisy signal whose STFT modulus is represented in (a), along with the true CR, for  $t$  between 0.3 and 0.5; (e): same as (d) but for the signal whose STFT modulus is represented in (b); (f): same as (d) but for the signal whose STFT modulus is represented in (c). The input SNR is 10 dB.

following first order Taylor expansion,

$$\begin{aligned}\bar{q}_{f+\varepsilon} &= -\frac{1}{2\pi} \Im \left\{ \frac{V_f^g + V_\varepsilon^g}{V_f^{t^2g} + V_\varepsilon^{t^2g}} \right\} = -\frac{1}{2\pi} \Im \left\{ \frac{V_f^g}{V_f^{t^2g}} \frac{1}{1 + \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}}} + \frac{V_\varepsilon^g}{V_f^{t^2g}} \frac{1}{1 + \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}}} \right\} \\ &\approx -\frac{1}{2\pi} \Im \left\{ \frac{V_f^g}{V_f^{t^2g}} \left(1 - \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}}\right) + \frac{V_\varepsilon^g}{V_f^{t^2g}} \left(1 - \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}}\right) \right\} \approx \bar{q}_f + \frac{1}{2\pi} \Im \left\{ \frac{V_f^g V_\varepsilon^{t^2g}}{(V_f^{t^2g})^2} - \frac{V_\varepsilon^g}{V_f^{t^2g}} \right\}.\end{aligned}\quad (5.11)$$

In this last expression, as  $\bar{q}_f(t, \eta)$  is very close to  $\phi''(t)$ , this means that the other term is responsible for the estimation error, that we detail hereafter,

$$\begin{aligned}\Im \left\{ \frac{V_f^g V_\varepsilon^{t^2g}}{(V_f^{t^2g})^2} - \frac{V_\varepsilon^g}{V_f^{t^2g}} \right\} &= \Im \left\{ \frac{V_f^g}{V_f^{t^2g}} \left( \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}} - \frac{V_\varepsilon^g}{V_f^g} \right) \right\} \\ &= \Im \left\{ \frac{V_f^g}{V_f^{t^2g}} \right\} \Re \left\{ \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}} - \frac{V_\varepsilon^g}{V_f^g} \right\} \\ &\quad + \Re \left\{ \frac{V_f^g}{V_f^{t^2g}} \right\} \Im \left\{ \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}} - \frac{V_\varepsilon^g}{V_f^g} \right\} \\ &= -2\pi \phi''(t) \Re \left\{ \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}} - \frac{V_\varepsilon^g}{V_f^g} \right\} + \frac{2\pi}{\sigma^2} \Im \left\{ \frac{V_\varepsilon^{t^2g}}{V_f^{t^2g}} - \frac{V_\varepsilon^g}{V_f^g} \right\}.\end{aligned}\quad (5.12)$$

We therefore define the error function by,  $G(t) := \frac{V_\varepsilon^{t^2g}(t, \phi'(t))}{V_f^{t^2g}(t, \phi'(t))} - \frac{V_\varepsilon^g(t, \phi'(t))}{V_f^g(t, \phi'(t))}$ . As we saw previously on Figure 5.1 (d), (e) and (f),  $G$  is oscillatory, we study its *power spectral density* (PSD) for the sake of removing these oscillations.

### 5.0.2. Power Spectral Density of the Error

#### Case of a Pure Harmonic Signal

First, in the particular case where  $\phi''(t)$  is null, namely  $f(t) = e^{2i\pi\eta_0 t}$ , we remark that  $V_f^g(t, \phi'(t)) = f(t)$  and rewrite:

$$\begin{aligned}G(t) &= \frac{2\pi}{\sigma^2} \frac{V_\varepsilon^{t^2g}(t, \phi'(t))}{V_f^g(t, \phi'(t))} - \frac{V_\varepsilon^g(t, \phi'(t))}{V_f^g(t, \phi'(t))} \\ &= \left( \frac{2\pi}{\sigma^2} V_\varepsilon^{t^2g}(t, \phi'(t)) - V_\varepsilon^g(t, \phi'(t)) \right) e^{-2i\pi\eta_0 t},\end{aligned}\quad (5.13)$$

to get the following:

**Proposition 5.0.2.** *The power spectral density of  $G(t)$  is*

$$P_G(\eta) = \sigma_\varepsilon^2 \sigma^6 4\pi^2 \eta^4 e^{-2\pi\sigma^2\eta^2}.$$

The proof is available in Appendix B.2.

### Case of a Linear Chirp

Now let us consider the more general case of a linear chirp  $f(t) = e^{2i\pi(at + \frac{b}{2}t^2)} = e^{2i\pi\phi(t)}$ , for which we may write on the STFT ridge, using (5.8),

$$\frac{V_f^g}{V_f^{t^2g}} = \frac{2\pi(1 + (b\sigma^2)^2)}{\sigma^2(1 + ib\sigma^2)} = \frac{2\pi}{\sigma^2}(1 - ib\sigma^2), \quad (5.14)$$

and thus:

$$G(t) = \frac{\frac{2\pi}{\sigma^2}(1 - ib\sigma^2)V_\varepsilon^{t^2g}(t, \phi'(t)) - V_\varepsilon^g(t, \phi'(t))}{V_f^g(t, \phi'(t))}, \quad (5.15)$$

leading to the following:

**Proposition 5.0.3.** *The power spectral density of  $G(t)$  is*

$$P_G(\eta) = \frac{\sigma_\varepsilon^2 \sigma^6 4\pi^2 \eta^4}{(1 + b^2\sigma^4)^2} e^{-\frac{2\pi\sigma^2\eta^2}{1+b^2\sigma^4}}.$$

The proof is available in Appendix B.3.

### 5.0.3. Filtering the Oscillatory Error

Using the PSD computed above, we justify the use of a low-pass filter on  $\hat{q}_{\tilde{f}}$ . We first look for the maximum value of  $P_G$ , by focusing on  $H_G := \sqrt{P_G}$ , as it has its maximum at the same frequency and has a simpler expression. In the case of a PH, when  $\phi'' = 0$ , we remark that

$$H_G\left(\frac{1}{\sqrt{\pi}\sigma}\right) = 2\sigma\sigma_\varepsilon e^{-1}. \quad (5.16)$$

For the purpose of defining a low-pass filter, we first define a cutoff value that we denote  $\eta_{c,0}$ . We propose to choose  $c \in ]0, 1]$ , and set the respective fraction of the maximum value such that  $H_G(\eta_{c,0}) = cH_G\left(\frac{1}{\sqrt{\pi}\sigma}\right)$ . To solve such an equation, we make use of the Lambert W function, denoted by  $\mathcal{W}$ ,

$$\sigma_\varepsilon \sigma^3 2\pi \eta_{c,0}^2 e^{-\pi \sigma^2 \eta_{c,0}^2} = 2\sigma \sigma_\varepsilon c e^{-1} \Leftrightarrow \eta_{c,0} = \frac{\sqrt{-\mathcal{W}(-ce^{-1})}}{\sigma \sqrt{\pi}}. \quad (5.17)$$

In the case of a linear chirp, when  $\phi'' = b$  and with the same reasoning, the maximum is reached at  $\eta_m = \frac{\sqrt{1+b^2\sigma^4}}{\sigma \sqrt{\pi}}$ . It verifies

$$H_G(\eta_m) = 2\sigma_\varepsilon \sigma e^{-1}. \quad (5.18)$$

Then, similarly to the PH case, we set  $\eta_{c,b}$  at which  $H_G(\eta_{c,b}) = cH_G(\eta_m)$  and solve its corresponding equation,

$$\frac{\sigma_\varepsilon \sigma^3 2\pi \eta_{c,b}^2}{(1+b^2\sigma^4)} e^{-\frac{\pi \sigma^2 \eta_{c,b}^2}{1+b^2\sigma^4}} = 2c\sigma_\varepsilon \sigma e^{-1} \Leftrightarrow \eta_{c,b} = (1+\sigma^4 b^2)^{1/2} \frac{\sqrt{-\mathcal{W}(-ce^{-1})}}{\sigma \sqrt{\pi}}. \quad (5.19)$$

However, this time,  $\eta_{c,b}$  depends on the chirp rate  $b$ . Therefore, in the numerical simulations, we are going to set the low-pass filter at frequency  $\eta_{c,0}$ . As  $\eta_{c,0} \leq \eta_{c,b}$ , we expect to filter out more energy.

## 5.1. Evaluating the CR Filtering Process

In this section, we evaluate the quality of the filtered CR estimate that we denote by  $F(\hat{q}_{\hat{f}})$  with  $\hat{q}_{\hat{f}}$ . We make this comparison in terms of output SNR with the true CR, namely  $SNR(\phi'', E - \phi'')$ , where  $E$  is any of the estimates. We also set  $c = \frac{1}{\sqrt{10}}$  resulting on a cut-off frequency at 10% of the maximum value. We study this choice of  $c$  for modulated and non-modulated modes in order to draw conclusions.

We first illustrate the effect of the filter in Figure 5.2 (d), (e) and (f). We remark that most of the oscillations are removed even though some remain but at low amplitude. Then, in Figure 5.2 (a), (b) and (c), we notice that up to a SNR of 40 dB, for the three signals considered,  $F(\hat{q}_{\hat{f}})$  offers significant improvement. Indeed, a close look at the box-plots tells us that the worst estimation of  $F(\hat{q}_{\hat{f}})$  is significantly better than the best estimate of  $\hat{q}_{\hat{f}}$ , assuming that the RD is able to detect the ridge, which may not be the case for the signal of Figure 5.1 (c) at 0 dB.

The purpose of Figure 5.3 is to show that the threshold  $\eta_{c,0}$  is still interesting to consider, even when the signal is modulated. On Figure 5.3 (a), we remark that the same cut-off frequency can be used, resulting only in a small loss of energy. On Figure



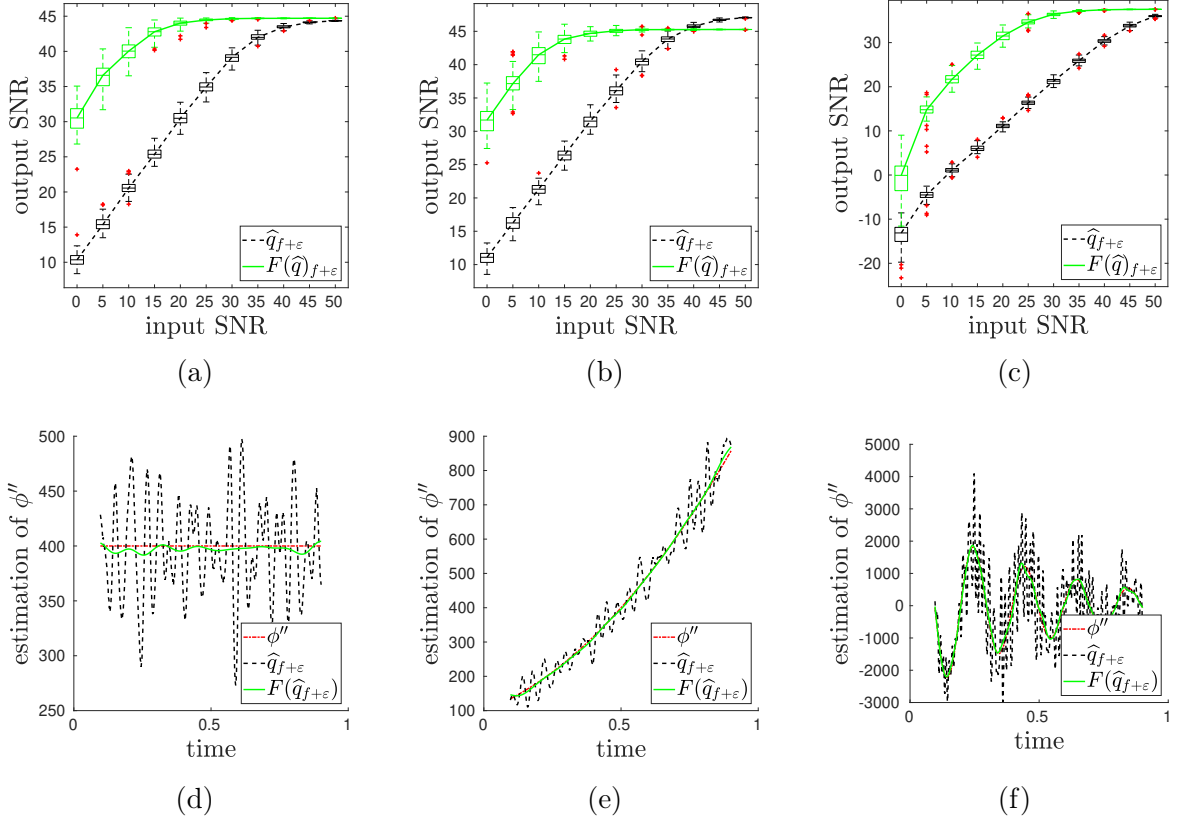


Figure 5.2.: (a): boxplot corresponding to the output SNR associated with CR estimators  $\hat{q}_{f+\varepsilon}$  or  $F(\hat{q}_{f+\varepsilon})$  evaluated on the STFT ridges of the noisy signal corresponding to Figure 5.1 (a) over 100 realizations of the noise when the input SNR varies; (b): same as (a) but for the noisy signal corresponding to Figure 5.1 (b); (c): same as (a) but for the noisy signal corresponding to Figure 5.1 (c); (d):  $\hat{q}_{f+\varepsilon}$  and  $F(\hat{q}_{f+\varepsilon})$  computed for a noisy version of the signal of Figure 5.1 (a), along with the ground truth  $\phi''(t)$  (input SNR = 10 dB); (e): same as (d) but for the signal of Figure 5.1 (b); (f): same as (d) but for the signal of Figure 5.1 (c).

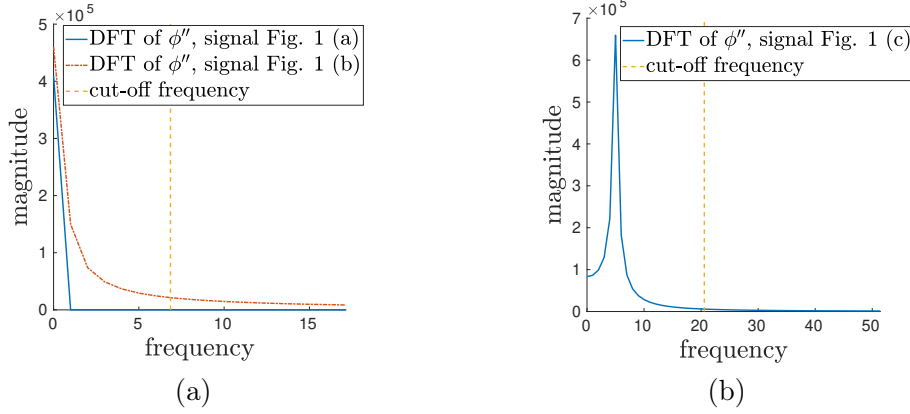


Figure 5.3.: (a): magnitude of the DFT of  $\phi''$  for the signals of Figure 5.1 (a) and Figure 5.1 (b); (b): same as (a) but for the signal of Figure 5.1 (c).

5.3 (b), for the most modulated signal, the loss of energy is even smaller.

### 5.1.1. Improving Linear Chirp Based Mode Retrieval

In this section, we investigate how to use the filtered CR estimator defined in Section 5.0.3 to improve LCR technique introduced in Section 2.6. We recall that this technique assumes a local linear chirp approximation for a mode, thus its STFT can be approximated by [19]:

$$V_f^g(t, \eta) \approx V_{f,approx}^g(t, \eta) := A(t) e^{-\pi\sigma^2 \frac{1+iC(t)\sigma^2}{1+(C(t)\sigma^2)^2} (\eta-B(t))^2}, \quad (5.20)$$

where  $A(t)$ ,  $B(t)$  and  $C(t)$  are respectively estimates of  $V_f^g(t, \phi'(t))$ ,  $\phi'(t)$  and  $\phi''(t)$ . In that context, mode reconstruction is performed by summing the approximation given by (5.20) over frequencies to obtain:

$$f(t) \approx \frac{1}{g(0)} \int_{\mathbb{R}} V_{f,approx}^g(t, \eta) d\eta, \quad (5.21)$$

LCR technique is then based on a specific choice for  $A(t)$ ,  $B(t)$  and  $C(t)$ . To compute the estimate  $B(t)$  of  $\phi'(t)$ , the technique uses the *local instantaneous frequency* estimator  $\hat{\omega}_f^{[2]} = \Re \left\{ \tilde{\omega}_f^{[2]} \right\}$  that we introduced with FSST2 in section 1.2, with:

$$\tilde{\omega}_f^{[2]} = \begin{cases} \tilde{\omega}_f + \frac{\partial_{\eta} \tilde{\omega}_f}{\partial_{\eta} \tilde{t}_f} \times (t - \tilde{t}_f) & \text{if } \partial_{\eta} \tilde{t}_f \neq 0 \\ \tilde{\omega}_f & \text{otherwise,} \end{cases} \quad (5.22)$$

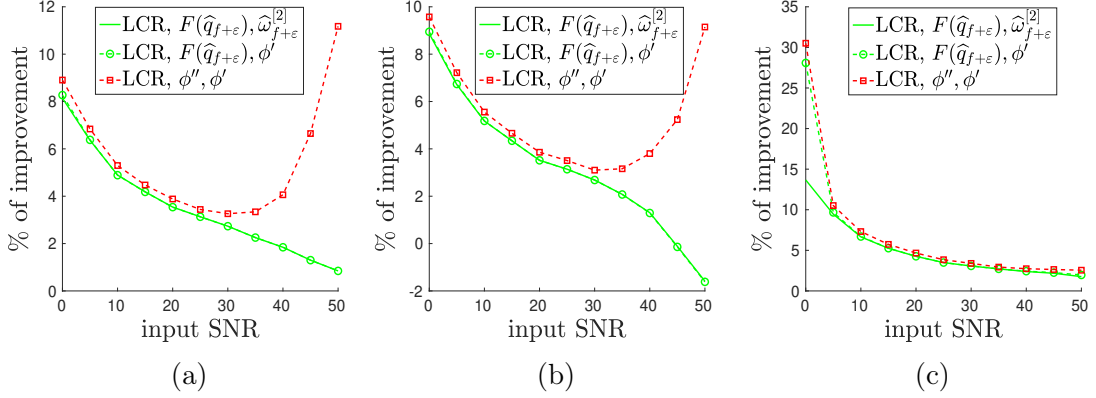


Figure 5.4.: (a): percentage of improvement for the LCR technique computed using (5.27) for the signal of Figure 5.1 (a) where the estimates for  $(\phi'', \phi')$  are  $(F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t))), \omega_{f+\varepsilon}^{[2]}(t, R_{f+\varepsilon}(t)))$ ,  $(F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t))), \phi')$ , or  $(\phi'', \phi')$ ; (b): same as (a) but for the signal of Figure 5.1 (b); (c): same as (a) but for the signal of Figure 5.1 (c)

where  $\tilde{\omega}_f$  and  $\tilde{t}_f$  were introduced in Section 1.2. More precisely, denoting by  $R_f(t)$  the STFT ridge corresponding to  $f$ ,  $B(t)$  is set to  $\hat{\omega}_f^{[2]}(t, R_f(t))$ ,  $C(t)$  to  $\hat{q}_f(t, R_f(t))$ , and finally  $A(t)$  to  $V_f^g(t, \hat{\omega}_f^{[2]}(t, R_f(t)))$ . When one considers the noisy signal  $f + \varepsilon$  with  $n$  a complex Gaussian white noise, similar estimates are derived replacing  $f$  by  $f + \varepsilon$  in the definitions of  $A$ ,  $B$  and  $C$ , which are thus all impacted by noise.

We now investigate the improvement brought by setting  $C(t)$  to  $F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t)))$  instead of  $\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t))$  in LCR technique,  $A(t)$  and  $B(t)$  remaining unchanged. We are going to compare this technique to original LCR and, to measure the impact of the estimator  $B$  on LCR technique, to implementations of LCR in which  $B(t)$  is set to  $\phi'(t)$  and  $C(t)$  either to  $F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t)))$  or to  $\phi''(t)$ . To clarify this, as LCR can be viewed as a function of 3 parameters, we define:

$$LCR_{A,B,C}(t, \eta) := A(t) e^{-\pi \sigma^2 \frac{1+iC(t)\sigma^2}{1+(C(t)\sigma^2)^2} (\eta - B(t))^2}. \quad (5.23)$$

In that framework, the original LCR technique corresponds to

$$(A, B, C) = (A_0, B_0, C_0) := \left( V_{f+\varepsilon}^g(t, \hat{\omega}_{f+\varepsilon}^{[2]}(t, R_{f+\varepsilon}(t))), \right. \\ \left. \hat{\omega}_{f+\varepsilon}^{[2]}(t, R_{f+\varepsilon}(t)), \hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t)) \right), \quad (5.24)$$

the proposed technique to:

$$(A, B, C) = (A_0, B_0, F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t)))), \quad (5.25)$$

while the other two references to

$$\begin{aligned} (A, B, C) &= (A_0, \phi'(t), F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t)))) \\ (A, B, C) &= (A_0, \phi'(t), \phi''(t)). \end{aligned} \quad (5.26)$$

For a monocomponent signal  $f$ , we define  $f_{(A,B,C)}$  the reconstructed signal using LCR technique with the set of parameters  $(A, B, C)$ . Then, using the definition of  $SNR_{out}$  introduced in (2.21, page 22), we compute the percentage of improvement brought by using a set  $(A, B, C)$  different from  $(A_0, B_0, C_0)$  as follows:

$$\frac{SNR_{out}(f_{(A,B,C)}, f) - SNR_{out}(f_{(A_0,B_0,C_0)}, f)}{SNR_{out}(f_{(A_0,B_0,C_0)}, f)} \times 100. \quad (5.27)$$

The results are depicted in Figure 5.4 for the three signals of Figure 5.1 and when  $(A, B, C)$  are either defined by (5.25) or (5.26). First, these clearly show the improvement brought by using  $F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t)))$  instead of  $\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t))$ . Second, these results also tell us that to know the actual  $\phi'(t)$  does not result in an extra improvement when compared with the proposed method. Finally, we notice that to know the actual  $\phi''$  enables a better mode reconstruction, but only at high SNRs and only for the first two signals. On the contrary, the gain of knowing the true IF and CR does not result in a significant improvement in terms of reconstruction for the last signal, except at very low SNR, but then ridge detection may fail, making the interpretation of the results more difficult.

So, we reach the conclusion that the proposed estimator  $(F(\hat{q}_{f+\varepsilon}(t, R_{f+\varepsilon}(t))), \hat{\omega}_{f+\varepsilon}^{[2]}(t, R_{f+\varepsilon}(t)))$  for  $(\phi'', \phi')$  leads to nearly optimal reconstruction. If one would like to further improve mode reconstruction with LCR technique one should investigate how to find a better estimate of  $V_f^g(t, \phi'(t))$  than  $A_0(t)$ , but this is beyond the scope of this thesis.

## 5.2. Frequency Variation of a Voice Signals

In this section, the objective is to study the stability of the estimate  $F(\hat{q}_{\hat{f}})$  which we apply to the same voice signal as in Section 3.4.3, namely the recording 62 of the Saarbrücken Voice Database [65]. The signal depicted in Figure 5.5 shows its TFR with superimposed ridge. We recall that it contains sustained /a/ vowels of the type “low-high-low”, and

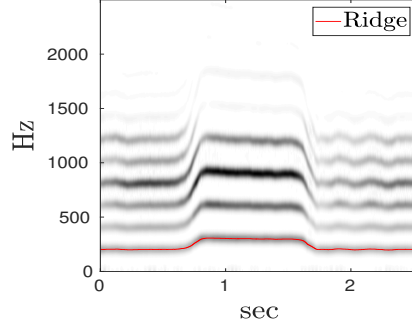


Figure 5.5.: ridge corresponding to the fundamental frequency superimposed on the STFT of the recording 62 taken from [65].

we want our CR estimate to preserve the change in pitch. On figure 5.6, we compare the stability of  $F(\hat{q}_{\hat{f}})$  and  $\hat{q}_{\hat{f}}$  over 100 realizations of noise. We test this for three noise levels, in 5.6 (a), (d) for 5dB, 5.6 (b), (e) for 10dB and 5.6 (c), (f) for 15dB. In all cases, the 5% and 95% quantiles associated with the estimation of  $F(\hat{q}_{\hat{f}})$  are much closer to the estimation without noise and still preserve the variation of the pitch that we remark on Figure 5.5 around 0.75s and 1.65s. We may therefore conclude that  $F(\hat{q}_{\hat{f}})$  seems to be improved in terms of stability with respect to  $\hat{q}_{\hat{f}}$ .

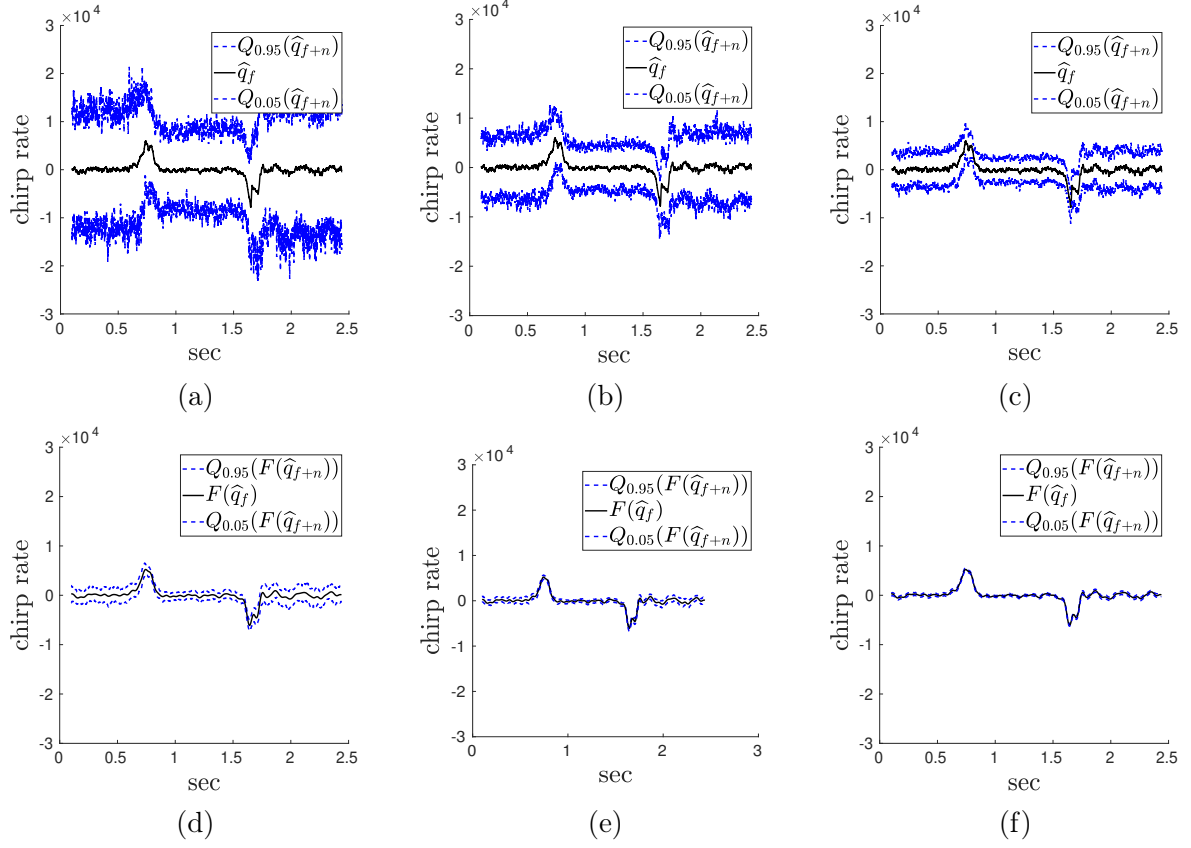


Figure 5.6.: (a): for 100 realizations of the noise, the range corresponding to the 5% and 95% quantiles of  $\hat{q}_{f+\varepsilon}$  evaluated on the STFT ridge corresponding to the fundamental frequency of the noisy signal whose STFT modulus is displayed in Figure 5.5,  $\hat{q}_f$  on the ridge corresponding to the fundamental frequency of the noiseless signal being also superimposed (input SNR = 5 dB for the noise); (b): same as (a) but with input SNR= 10 dB; (c): same as (a) but with input SNR= 15 dB. (d): same as (a) but using the filtered estimator  $F(\hat{q}_{f+\varepsilon})$  instead of  $\hat{q}_{f+\varepsilon}$ ,  $F(\hat{q}_f)$  is also superimposed ; (e): Same as (d) but with input SNR= 10 dB; (f): same as (d) but with input SNR= 15 dB.

### 5.3. Conclusion

In this chapter, we introduced the filtering of the CR estimator associated with FSST2 to remove most of the oscillatory errors due to noise. We carried out the theoretical study of the estimator in both pure harmonic and linear chirp cases. We also evaluated the filtered estimate on numerical experiments, first on synthetic signals, where the filtered estimator gives significant improvement compared to the original one, in terms of output SNR. In a second part, we verified on a real voice signal that it is also more robust, meaning that the noise has a smaller effect on the estimation. From these results, we conclude that the filtered estimate is significantly more robust to noise.

## 6. SST Based Heart Rate Estimation

### 6.1. Introduction

The estimation of the HR is still an active topic of research and includes various approaches such as time based R-peak detection [79, 80] or TF analysis [81, 82]. The latter is relevant for the study of physiological signals, as they can be highly non-stationary. ECG signals are such cases, in which the shape of heart beat, namely the QRS complex, varies and can make time based analysis difficult. ECGs are actively studied using TF analysis, for instance, it is used to detect sleep apnea [83], to study arrhythmia [84], to identify coronary artery diseases [85] or to separate fetal ECG from that of the mother on abdominal ECG signal of pregnant women [86].

In this chapter, our goal is to estimate the HR of an ECG signal, which we assume to be harmonic. Thus, we are going to look for the *fundamental frequency* (FF) as it represents an estimate of the HR. To perform this estimation, some existing TF techniques try to estimate the spectrogram using non-negative matrix factorization, as for instance on PCGs [87] and ECGs [88] signals. In our case, we propose to define a dictionary matching both the choice of TFR and the harmonicity of the signal, which we only use to estimate the HR.

The layout of the chapter is as follows: first we define the harmonic model of a MCS and then use *earth mover distance* (EMD) to select a component in a dictionary for HR estimation. Finally, we discuss the influence of the choice of TFR on the quality of HR estimation and perform numerical assessment of the technique on synthetic and real ECG signals.

---

This chapter is based on the work published in [44].



## 6.2. Sparsity of Harmonic Signals and Application to ECG Signals

A harmonic signal is a specific type of MCS defined by,

$$f(t) = \sum_{p=1}^P A_p(t) e^{2i\pi p\phi(t)}. \quad (6.1)$$

For such a signal,  $\phi'(t)$  is the FF. In the case of an ECG signal, the instantaneous HR in *beats per minute* (bpm) at time  $t$  is given by  $60\phi'(t)$ .

### 6.2.1. Dictionary Based Analysis of the Magnitude

We propose to study the magnitude of harmonic signals in a TFR  $\mathbf{X}$  using a dictionary  $\mathbf{W}$  containing  $K_e$  components. A good choice of the number of components  $K_e$  enables a good frequency resolution. In order to compare  $\mathbf{X}$  with a component of  $\mathbf{W}$ , we propose to use the EMD and to collect this information in a matrix  $\mathbf{H}$ . We finally use the latter to select the best component and estimate the heart rate. The dictionary matrix  $\mathbf{W}$  is fixed and describes possible structures of harmonicity depending on the variation of the FF. In the case of ECG signals, we consider a variation of the FF between 30 and 180 bpm and thus, with an accuracy of 1 bpm, set  $K_e$  to 151.

In Figures 6.1 (a) and (b), we respectively display the STFT and FSST2 magnitude of one real ECG signal. We look for dictionaries allowing us to reasonably characterize these TFRs, and we assume FSST and FSST2 share the same dictionary. In the case of STFT, the dictionary takes into account the spreading of the energy in the TF plane. We denote this dictionary by  $\mathbf{W}^0$  and give an illustration in Figure 6.1 (c), where the spreading of energy assumes that the analysis window is Gaussian. In the case of FSST and FSST2, the dictionary is built assuming perfect energy reassignment, meaning a single nonzero coefficient is enough to represent a harmonic. We denote this dictionary by  $\mathbf{W}^1$  and give an illustration in Figure 6.1 (d). Note that for these two dictionaries, a high component index takes more harmonics into account than for a low component index. We assume that, as long as the highest component has enough harmonics and that the frequency range of the dictionary takes the sampling rate into account, it provides a good representation of the signal. We can confirm this in Figures 6.1 (a) and (b), where most of the energy is located in the lower part of the TFR.

Our objective is to compare HR detection when the considered representations are STFT, FSST and FSST2. We do not take into account higher orders of FSSTN as it

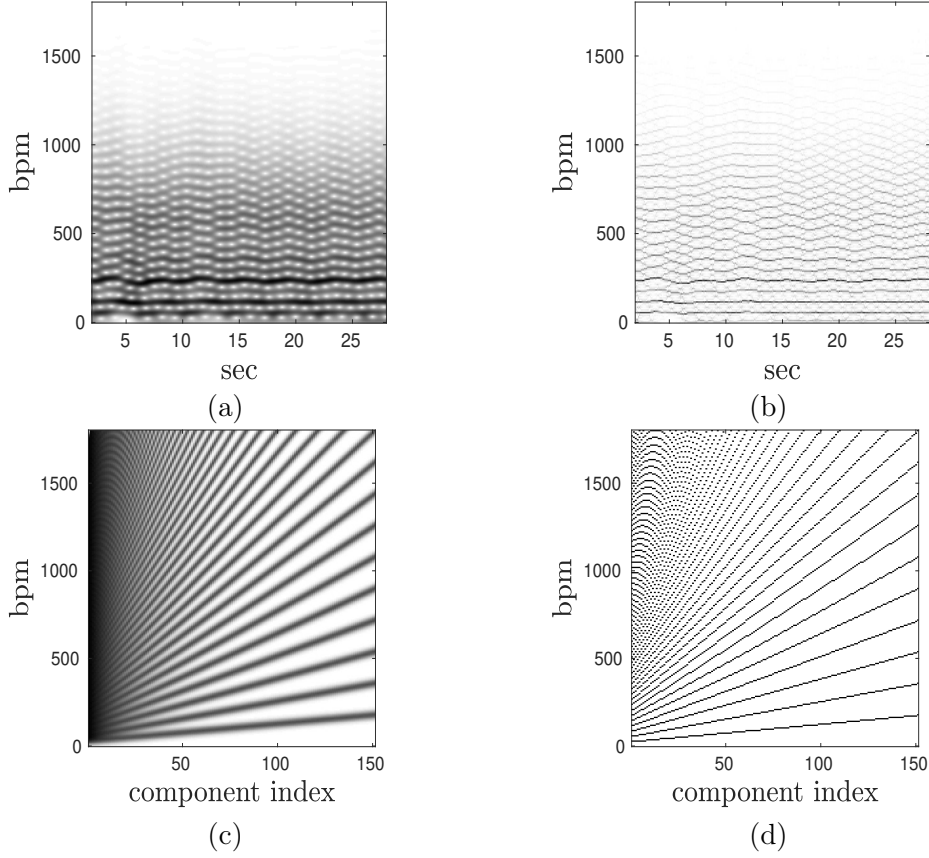


Figure 6.1.: (a): STFT of an ECG generated following [92]; (b): FSST2 of the same ECG; (c): Dictionary matrix  $\mathbf{W}^1$  used with STFT-based approach; (d): Dictionary matrix  $\mathbf{W}^0$  used with FSST2-based approach. For the sake of legibility, in (c) and (d) we do not normalize the column of the matrices.

does not increase the accuracy of the estimation.

In order to make a decomposition, one needs to set the cost function  $D$ . In the case where  $D$  is the Euclidian distance, the decomposition is a least-square minimization problem,

$$\mathbf{H} = \underset{\mathbf{H}_0}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{W}\mathbf{H}_0\|^2,$$

and thus  $\mathbf{H}$  satisfies  $\mathbf{W}^T\mathbf{W}\mathbf{H} = \mathbf{W}^T\mathbf{X}$ , and is not necessarily unique because  $\mathbf{W}^T\mathbf{W}$  is not invertible. Since the FF is associated with only one component of  $\mathbf{W}$ , we impose stronger conditions to obtain the unicity of the solution. At each time instant, we compute the EMD [93] between the  $i$ -th column of  $\mathbf{W}$  and the  $n$ -th column of  $\mathbf{X}$  to obtain

a matrix  $\mathbf{H}$  defined by,

$$\mathbf{H}_{i,n} = d(\mathbf{W}_i, \mathbf{X}_n), \quad (6.2)$$

where  $d$  is the EMD.

In this setting, finding a minimum in  $\mathbf{H}$  is equivalent to finding the most relevant component to associate with the signal at a time index, and to estimate the associated FF. We denote this component by  $i_0$  which satisfies,

$$i_0 = \underset{i}{\operatorname{argmin}} \mathbf{H}_{i,n}. \quad (6.3)$$

In the following section, we propose an algorithm to estimate the HR using  $i_0$  for which we take into account the accuracy induced by the choice of the representation for  $\mathbf{X}$ .

### 6.2.2. Algorithm for HR Computation

We consider the HR estimate as the minimum on  $i$  of  $\underset{i}{\operatorname{argmin}} d(\mathbf{W}_i, \mathbf{X}_n)$ . To assess the quality of this estimate, we set

$$\hat{i}_M = \operatorname{median}_n \left( \underset{i \in 1, \dots, K_e}{\operatorname{argmin}} d(\mathbf{W}_i, \mathbf{X}_n) \right) \quad (6.4)$$

and consider it as an estimate of the mean HR. We now seek to evaluate the quality of the HR estimate given by the minimum of the EMD. For that purpose, we define the following range of frequencies,  $\hat{I} := [\frac{3\hat{i}_M}{4}, \frac{3\hat{i}_M}{2}]$  that we consider as valid, and then define  $\hat{B}$  as follows:

$$\hat{B} := \frac{\#\{n, \underset{i}{\operatorname{argmin}} d(\mathbf{W}_i, \mathbf{X}_n) \in \hat{I}\}}{N}. \quad (6.5)$$

The choice for  $\hat{I}$  is supported by the fact that, assuming that the HR varies slowly, the upper harmonic admits  $\frac{3\hat{i}_M}{2}$  as a lower bound and that no minimum should be detected under the lower bound of the FF, namely  $\frac{3\hat{i}_M}{4}$ . Based on this model, we may define the probability of false detection as  $P_f = 1 - \hat{B}$ .

However, in order to trust the median, we have to make sure that  $P_f < \frac{1}{2}$  and that the signal is sufficiently long. For that purpose, imposing that the median of the detection computed from time indices  $1, \dots, n_{med}$  is exact with probability  $p_0$  or above, we compute

the minimal value of  $n_{med}$  using

$$n_{med} = \min \left( n \geq 1 \text{ s.t. } \sum_{k=0}^{n/2} C_n^k P_f^k (1 - P_f)^{n-k} \leq p_0 \right). \quad (6.6)$$

The estimation of the HR at time indices  $1, \dots, n_{med}$  can then be done through Algorithm 2, where we look for the components in  $\hat{I}$  to avoid selecting a component related to noise or harmonics. Then, having defined the HR for the first  $n_{med}$  indices, the estimation carries on with Algorithm 3 which is used to estimate HR after time instant  $n_{med}$ . To make our detection more robust, we propose to take into account the standard deviation of previous detections, denoted by  $\Delta_n$ , to define a range of relevant components. Such a range is defined using a parameter  $\gamma$ , the choice of which is discussed in Section 6.2.4.

Note that because Algorithm 3 considers only a subset of  $H$ , it does not require to compute the EMD for all components  $K_e$ .

---

**Algorithm 2:** HR estimation step 1

---

**Input** :  $\mathbf{W}, \mathbf{X}, n_{med}$   
**for**  $n = 1, \dots, n_{med}$  **do**  
     $\tilde{\mathbf{i}}[n] = \underset{i \in [1, K_e]}{\operatorname{argmin}} d(\mathbf{W}_i, \mathbf{X}_n)$   
**end**  
 $\hat{i}_M = \operatorname{median} \left\{ \tilde{\mathbf{i}}[n], n = 1, \dots, n_{med} \right\}$   
**for**  $n = 1, \dots, n_{med}$  **do**  
     $\hat{\mathbf{i}}[n] = \underset{i \in [\frac{3}{4}\hat{i}_M, \frac{3}{2}\hat{i}_M]}{\operatorname{argmin}} d(\mathbf{W}_i, \mathbf{X}_n)$   
**end**

---



---

**Algorithm 3:** HR estimation step 2

---

**Input** :  $\mathbf{W}, \mathbf{X}, n_{med}, \hat{\mathbf{i}}[1, \dots, n_{med}], \gamma$   
**for**  $n = n_{med} + 1, \dots, L$  **do**  
     $\Delta_n = \operatorname{std} \left\{ \hat{\mathbf{i}}[k], k = 1, \dots, n - 1 \right\}$   
     $\hat{\mathbf{i}}[n] = \underset{[ \hat{i}[n-1] - \gamma \Delta_n, \hat{i}[n-1] + \gamma \Delta_n ]}{\operatorname{argmin}} d(\mathbf{W}_i, \mathbf{X}_n)$   
**end**

---

HR Mean	$\hat{B}$ STFT	$\hat{B}$ FSST	$\hat{B}$ FSST2
60 bpm	39.61%	<b>94.34%</b>	75.4%
70 bpm	81.18%	<b>99.14%</b>	90.04%
80 bpm	100%	100%	100%

Table 6.1.: Computation of  $\hat{B}$  on syntethic ECG signals with different HR means when either STFT, FSST or FSST2 are used to compute  $\mathbf{X}$

### 6.2.3. Determination of the Probability of False Detection

To compare the different probabilities of false detection depending on the type of TFR used to define  $\mathbf{X}$ , we consider synthetic ECGs generated as in [92], compute the corresponding STFT, FSST, and FSST2 and then (6.3) for each time instant when  $\mathbf{X}$  is associated with one of the just mentioned TFRs. With these synthetic signals, the mean HR component  $i_M$  is known, and we measure good detection by considering  $I = [\frac{3i_M}{4}, \frac{3i_M}{2}]$  and then

$$B = \frac{\#\{n, \underset{i}{\operatorname{argmin}} d(\mathbf{W}_i, \mathbf{X}_n) \in I\}}{N}, \quad (6.7)$$

the probability of false detection being  $1 - B$ .

We numerically notice that, with STFT or FSST2 the minimum of the EMD is much more likely to be associated with a sub-harmonic than when FSST is used, resulting in a much higher probability of false detection, and this is true whatever the mean HR ( the results are depicted in Table. 6.1).

An illustration of this is given in Figure 6.2 (a)-(c) when the TFR are either STFT, FSST and FSST2, respectively: the white curves correspond to the indices associated with the minimal EMD for each time index  $n$ . We notice that false detections usually occur when HR changes rapidly, and, as the signal is quasi-harmonic, to make a linear chirp approximation on the modes, as is done in FSST2, results in error at these time indices. As our plan is to use EMD for HR computation, we will make use of FSST rather than STFT or FSST2.

### 6.2.4. Study of Synthetic ECG

The goal of this section is to define an appropriate choice for parameter  $\gamma$  used in Algorithm 3. For that purpose, we investigate how good this algorithm is to estimate the

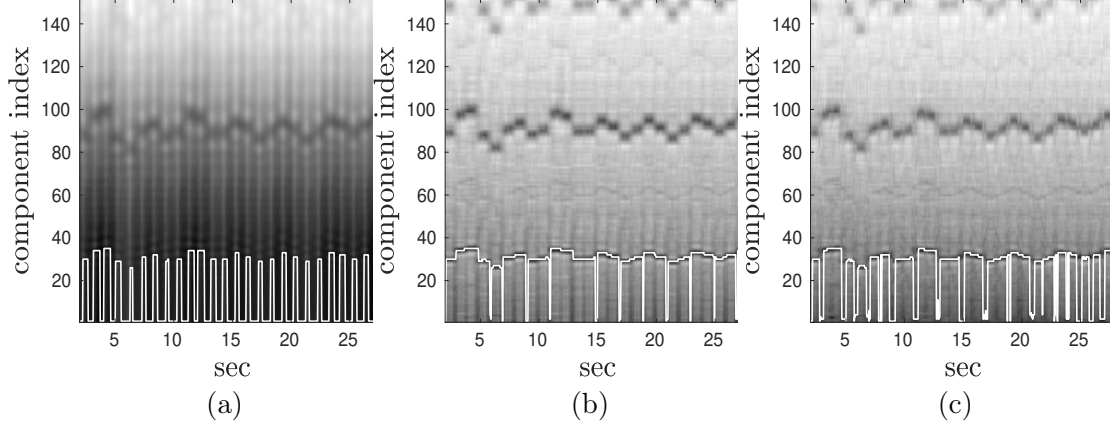


Figure 6.2.: In these figures, the HR varies around 65 bpm, which corresponds to the index 35 of the dictionary. (a): modulus of STFT, the white curve corresponds to the indices associated with the minimal EMD ( $\hat{B} = 39.61\%$ ); (b): same as (a) but for FSST ( $\hat{B} = 94.34\%$ ) (c): same as (a) but for FSST2 ( $\hat{B} = 75.4\%$ ).

mean and standard deviation of HR, which are input parameters of the function **ecgsyn** introduced in [92]. In what follows, we denote by  $HR_{syn}$  the mean heart rate and by  $\sigma_{syn}$  its standard deviation. To make HR estimation realistic, we consider  $HR_{syn} = 60$  bpm (corresponding to component index 31) and  $\sigma_{syn} = 2$  bpm. Such a signal enables a clear separation between harmonics and enough HR variability, so that its study is useful for the analysis of real ECGs that comes next. For that type of signals, we compute the mean of  $\hat{\mathbf{i}}$  along with its standard deviation, expressed in terms of a number of components, both as a function of  $\gamma$ . The results depicted in Figures 6.3 (a) and (b) show that the proposed algorithm enables very accurate estimations of both the mean and standard deviation of HR provided  $\gamma$  is taken around 2.5. Note that the range for  $\gamma$  to retrieve the mean and standard deviation of HR is very similar when either FSST or FSST2 are considered.

### 6.2.5. Study of Real ECGs

We here study ECGs extracted from SiSEC database [94], which correspond to thoracic recordings, which can be contaminated by different types of noises. Our HR estimation computed on such a signal, using  $\gamma = 2.5$ , fixed as explained in Section 6.2.4, is depicted in Figure 6.4 (a).

As FSST somehow reduces the noise while enhances the significant harmonics, the proposed HR estimation algorithm never gets trapped by minima that arise at sub-harmonic locations. We display HR estimations given by  $\hat{\mathbf{i}}$  in Figure 6.4 (b) (white line)

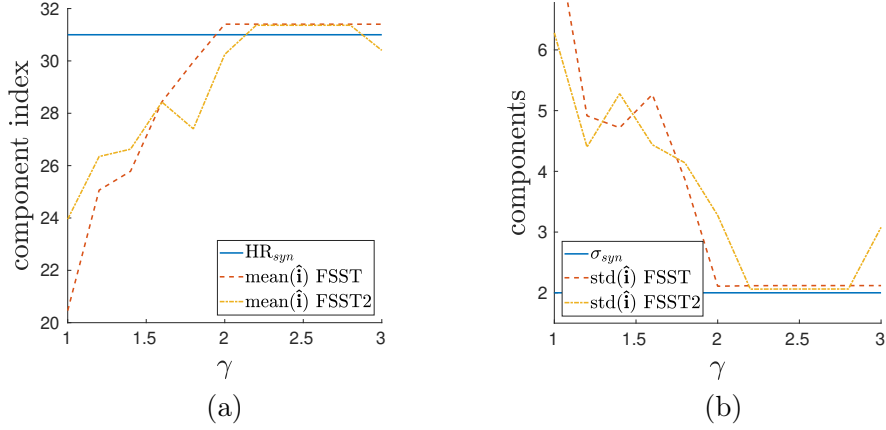


Figure 6.3.: (a): Estimation of HR mean from  $\hat{i}$ . (b): Estimation of the corresponding standard deviation.

along with the TF region corresponding to standard deviation  $\Delta_n$ , when  $n$  varies.

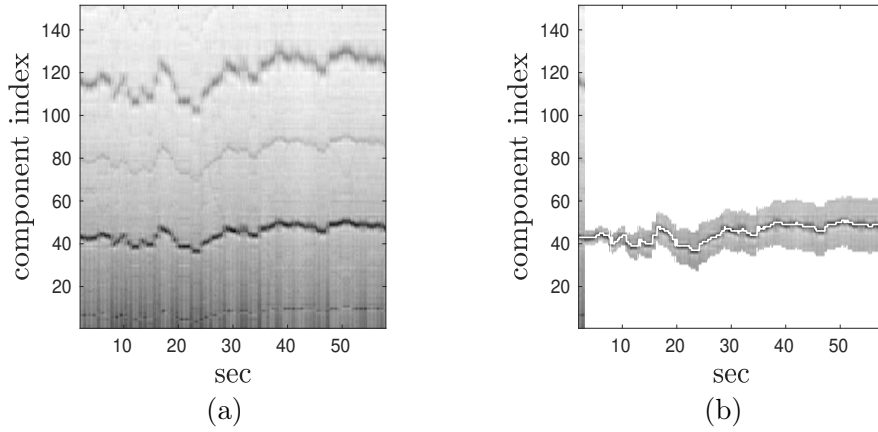


Figure 6.4.: (a) FSST based EMD of a real ECG signal. (b) HR detection on the same signal using  $\gamma = 2.5$ .

### 6.3. Conclusion

In this chapter, we studied the particular case of harmonic signals in the context of ECG to estimate the HR. We first showed that the choice of TFR can have a huge impact on the accuracy of estimations, and that even though FSST does not take frequency modulation into account, it provides the best results. Based on a reliable HR estimator and a relevant choice of TFR, we designed an algorithm to track the HR in time. We evaluated this algorithm on synthetic and real ECG signals to show numerically the

relevance of the proposed approach.





## 7. Conclusion

### 7.1. Contributions

In the context of this thesis, we performed STFT based TF analysis of MCSs.

In Chapter 2, the contribution is two-fold: first, we designed a technique to identify ridges in the TF plane under very noisy conditions. The key point was to take into account that there is no continuous ridge associated with a mode, and to gather ridge portions to identify each mode. The second contribution was to use a linear chirp model to perform MR. Given the ridges of the signals, we locally approximate the linear chirp and recover the TFR for each mode.

In Chapter 3, we defined interference structures called TFBs and performed their complete mathematical characterization for the case of two parallel linear chirps. For practical purposes, we designed a technique to detect the TFBs, based on the detection of merging ridges. Finally, we considered perspectives, using a different approach based on the spectrogram zeros.

In Chapter 4, we studied FSSTN and remarked that the selection of the order of this technique can be done by maximizing the concentration of energy. Then, since FSST2 performs well in noisy situations, we studied its associated CR estimator and observed that the errors due to noise are oscillatory. Therefore, we analyzed the power spectral density of the latter, to design a filtering technique, improving its accuracy.

In Chapter 6, we performed HR estimation on ECG signals, assuming that they can be modeled as a harmonic signal. We first compared the accuracy of a HR estimator based on the EMD, and demonstrated that the choice of the TFR can be critical. We finally designed an algorithm to detect the HR, taking into account the modulation of the signal.

### 7.2. Future research

To improve the state of the art, related to the techniques introduced in this thesis, we propose several directions we detail below.

Ridge detection in noisy conditions does not work on a general MCS. It can be improved by taking into account interference of crossing modes, and define ridges only when relevant, for example, when modes are partially defined in time.

The IF estimator associated with FSSTN is sensitive to noise and its improvement in terms of accuracy would make FSSTN more stable. In Chapter 4, we made progress towards such an estimate in the case of FSST2, by remarking that the nature of the error associated with the CR estimator is oscillatory, enabling the design of a denoising technique. Further research on high order IF estimators may lead to the design of an adaptive technique in noisy case, which should match the order of FSSTN with the phase for each mode of the signal.

Concerning the particular case of ECG signals, more knowledge on interference and TF localization can also help to identify the HR. Particularly, the case of measuring the fetal HR from a single abdominal ECG channel is a challenging task, since mother and fetal heart beats can interfere. Moreover, it also has the added difficulty that the fetal signal have less power and its HR is therefore more difficult to read.

# Bibliography

- [1] L. Cohen, *Time-frequency Analysis: Theory and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995, ISBN: 0-13-594532-1.
- [2] J. B. J. Baron Fourier, *The analytical theory of heat*. The University Press, 1878.
- [3] D. Gabor, “Theory of communication. part 1: The analysis of information”, *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, vol. 93, no. 26, pp. 429–441, 1946. DOI: [10.1049/ji-3-2.1946.0074](https://doi.org/10.1049/ji-3-2.1946.0074).
- [4] A. Grossmann and J. Morlet, “Decomposition of Hardy functions into square integrable wavelets of constant shape”, *SIAM journal on mathematical analysis*, vol. 15, no. 4, pp. 723–736, 1984.
- [5] S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989, ISSN: 0162-8828. DOI: [10.1109/34.192463](https://doi.org/10.1109/34.192463).
- [6] H Weyl, *The theory of groups and quantum mechanics, 1931, rept*, 1950.
- [7] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd. Academic Press, 2008, ISBN: 0123743702, 9780123743701.
- [8] J. Ville, “Theorie et application dela notion de signal analytique”, *Câbles et transmissions*, vol. 2, no. 1, pp. 61–74, 1948.
- [9] W. Martin and P. Flandrin, “Wigner-ville spectral analysis of nonstationary processes”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, pp. 1461–1470, 1985.
- [10] K. Kodera, C. De Villedary, and R. Gendrin, “A new method for the numerical analysis of non-stationary signals”, *Physics of the Earth and Planetary Interiors*, vol. 12, no. 2-3, pp. 142–150, 1976.
- [11] I. Daubechies and S. Maes, “A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models”, *Wavelets in medicine and biology*, pp. 527–546, 1996.

- [12] F. Hlawatsch and F. Auger, *Time-frequency analysis*. John Wiley & Sons, 2013.
- [13] M. Portnoff, “Time-frequency representation of digital signals and systems based on short-time fourier analysis”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 55–69, 1980.
- [14] R Crochiere, “A weighted overlap-add method of short-time fourier analysis/synthesis”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 99–102, 1980.
- [15] 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.
- [16] F. Auger and P. Flandrin, “Improving the readability of time-frequency and time-scale representations by the reassignment method”, *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, 1995. DOI: [10.1109/78.382394](https://doi.org/10.1109/78.382394).
- [17] T. Oberlin, S. Meignen, and V. Perrier, “Second-order synchrosqueezing transform or invertible reassignment? Towards ideal time-frequency representations”, *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1335–1344, 2015, ISSN: 1053-587X. DOI: [10.1109/TSP.2015.2391077](https://doi.org/10.1109/TSP.2015.2391077).
- [18] F. Auger *et al.*, “Time-frequency reassignment and synchrosqueezing: An overview”, *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 32–41, 2013. DOI: [10.1109/MSP.2013.2265316](https://doi.org/10.1109/MSP.2013.2265316).
- [19] R. Behera, S. Meignen, and T. Oberlin, “Theoretical analysis of the second-order synchrosqueezing transform”, *Applied and Computational Harmonic Analysis*, vol. 45, no. 2, pp. 379–404, 2018.
- [20] S. Meignen, T. Oberlin, and D.-H. Pham, “Synchrosqueezing transforms: From low-to high-frequency modulations and perspectives”, *Comptes Rendus Physique*, vol. 20, no. 5, pp. 449–460, 2019.
- [21] I. Daubechies, J. Lu, and H.-T. Wu, “Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool”, *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243–261, 2011.
- [22] R. Gribonval and E. Bacry, “Harmonic decomposition of audio signals with matching pursuit”, *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 101–111, 2003.

- 
- [23] M. Sandsten and J. Brynolfsson, "Classification of bird song syllables using wigner-ville ambiguity function cross-terms", in *2017 25th European Signal Processing Conference (EUSIPCO)*, IEEE, 2017, pp. 1739–1743.
  - [24] H.-I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 6, pp. 862–871, 1989.
  - [25] 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.
  - [26] U. R. Acharya, K. P. Joseph, N. Kannathal, L. C. Min, and J. S. Suri, "Heart rate variability", in *Advances in cardiac signal processing*, Springer, 2007, pp. 121–165.
  - [27] M. Malik and A. J. Camm, *Dynamic electrocardiography*. John Wiley & Sons, 2008.
  - [28] 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 journal of biomedical and health informatics*, vol. 21, no. 6, pp. 1533–1545, 2017.
  - [29] J. Bonnel, G. Le Touzé, B. Nicolas, and J. I. Mars, "Physics-based time-frequency representations for underwater acoustics: Power class utilization with waveguide-invariant approximation", *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 120–129, 2013.
  - [30] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank", *IEEE signal processing letters*, vol. 11, no. 2, pp. 112–114, 2004.
  - [31] G. Thakur and H.-T. Wu, "Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples.", *SIAM J. Math. Analysis*, vol. 43, no. 5, pp. 2078–2095, 2011.
  - [32] G. Thakur, E. Brevdo, N. S. Fuckar, and H.-T. Wu, "The synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications", *Signal Processing*, vol. 93, no. 5, pp. 1079–1094, May 2013, ISSN: 0165-1684. DOI: [10.1016/j.sigpro.2012.11.029](https://doi.org/10.1016/j.sigpro.2012.11.029).
  - [33] T. Oberlin, S. Meignen, and V. Perrier, "The Fourier-based synchrosqueezing transform", in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 315–319. DOI: [10.1109/ICASSP.2014.6853609](https://doi.org/10.1109/ICASSP.2014.6853609).

- [34] L. Li, H. Cai, and Q. Jiang, “Adaptive synchrosqueezing transform with a time-varying parameter for non-stationary signal separation”, *Applied and Computational Harmonic Analysis*, 2019.
- [35] L. Li, H. Cai, H. Han, Q. Jiang, and H. Ji, “Adaptive short-time Fourier transform and synchrosqueezing transform for non-stationary signal separation”, *Signal Processing*, vol. 166, p. 107 231, 2020.
- [36] C. K. Chui, Q. Jiang, L. Li, and J. Lu, “Analysis of an adaptive short-time fourier transform-based multicomponent signal separation method derived from linear chirp local approximation”, *Journal of Computational and Applied Mathematics*, p. 113 607, 2021.
- [37] S. Mann and S. Haykin, “The chirplet transform: Physical considerations”, *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2745–2761, 1995.
- [38] E. Candes and D. Donoho, “Ridglets: The key to high-dimensional intermittency”, *Phil. Trans. R. Soc. Lond*, pp. 2495–2509,
- [39] R. A. Carmona, W. L. Hwang, and B. Torr  sani, “Multiridge detection and time-frequency reconstruction”, *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 480–492, 1999.
- [40] 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.
- [41] N. Laurent and S. Meignen, “A novel time-frequency technique for mode retrieval based on linear chirp approximation”, *IEEE Signal Processing Letters*, vol. 27, pp. 935–339, 2020.
- [42] N. Laurent and S. Meignen, “A new adaptive technique for multicomponent signals reassignment based on synchrosqueezing transform”, in *2022 30th European Signal Processing Conference (EUSIPCO)*, IEEE, 2022, to appear.
- [43] N. Laurent, M. A. Colominas, and S. Meignen, “On local chirp rate estimation in noisy multicomponent signals: With an application to mode reconstruction”, *IEEE Transactions on Signal Processing*, vol. 70, pp. 3429–3440, 2022.
- [44] N. Laurent, S. Meignen, J. Fontecave-Jallon, and B. Rivet, “A novel algorithm for heart rate estimation based on synchrosqueezing transform”, in *2021 29th European Signal Processing Conference (EUSIPCO)*, IEEE, 2021, pp. 1286–1290.

- 
- [45] L. Stanković, “A measure of some time–frequency distributions concentration”, *Signal Processing*, vol. 81, no. 3, pp. 621–631, 2001.
  - [46] L. Stankovic, M. Dakovic, and V. Ivanovic, “Performance of spectrogram as IF estimator”, *Electronics Letters*, vol. 37, no. 12, pp. 797–799, 2001.
  - [47] R. Carmona, W. Hwang, and B. Torresani, “Characterization of signals by the ridges of their wavelet transforms”, *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2586–2590, 1997, ISSN: 1053-587X. DOI: [10.1109/78.640725](https://doi.org/10.1109/78.640725).
  - [48] I. Djurović and L. Stanković, “An algorithm for the Wigner distribution based instantaneous frequency estimation in a high noise environment”, *Signal Processing*, vol. 84, no. 3, pp. 631–643, 2004.
  - [49] M. A. Colominas, S. Meignen, and D.-H. Pham, “Fully adaptive ridge detection based on STFT phase information”, *IEEE Signal Processing Letters*, 2020.
  - [50] R. Carmona, W. Hwang, and B. Torresani, “Multiridge detection and time-frequency reconstruction”, *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 480–492, 1999, ISSN: 1053-587X. DOI: [10.1109/78.740131](https://doi.org/10.1109/78.740131).
  - [51] X. Zhu, Z. Zhang, J. Gao, and W. Li, “Two robust approaches to multicomponent signal reconstruction from STFT ridges”, *Mechanical Systems and Signal Processing*, vol. 115, pp. 720–735, 2019.
  - [52] S. Meignen, D.-H. Pham, and S. McLaughlin, “On demodulation, ridge detection, and synchrosqueezing for multicomponent signals”, *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2093–2103, 2017.
  - [53] D.-H. Pham and S. Meignen, “A novel thresholding technique for the denoising of multicomponent signals”, in *43th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2018.
  - [54] D. Donoho and I. Johnstone, “Ideal spatial adaptation via wavelet shrinkage”, *Biometrika*, vol. 81, pp. 425–455, 1994.
  - [55] L. Sachs, *Applied Statistics: A Handbook of Techniques*. New York: Springer-Verlag, 1984, vol. p.253.
  - [56] R. G. Baraniuk, P. Flandrin, A. J. Janssen, and O. J. Michel, “Measuring time-frequency information content using the Rényi entropies”, *IEEE Transactions on Information theory*, vol. 47, no. 4, pp. 1391–1409, 2001.



- [57] S. Meignen, M. Colominas, and D.-H. Pham, “On the use of Rényi entropy for optimal window size computation in the short-time Fourier transform”, in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, pp. 5830–5834.
- [58] D.-H. Pham and S. Meignen, “A novel thresholding technique for the denoising of multicomponent signals”, in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018, pp. 4004–4008.
- [59] J. Zheng, J. Cheng, and Y. Yang, “Partly ensemble empirical mode decomposition: An improved noise-assisted method for eliminating mode mixing”, *Signal Processing*, vol. 96, pp. 362–374, 2014.
- [60] B. P. Abbott *et al.*, “Observation of gravitational waves from a binary black hole merger”, *Physical review letters*, vol. 116, no. 6, p. 061 102, 2016.
- [61] D. H. Pham and S. Meignen, “High-order synchrosqueezing transform for multicomponent signals analysis-with an application to gravitational-wave signal.”, *IEEE Trans. Signal Processing*, vol. 65, no. 12, pp. 3168–3178, 2017.
- [62] B. P. Abbott, R. Abbott, *et al.*, “GW151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence”, *Physical Review Letters*, vol. 116, no. 24, p. 241 103, 2016.
- [63] S. Meignen, D.-H. Pham, and M. A. Colominas, “On the use of short-time fourier transform and synchrosqueezing-based demodulation for the retrieval of the modes of multicomponent signals”, *Signal Processing*, vol. 178, p. 107 760, 2021.
- [64] 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.
- [65] B. Woldert-Jokisz. “Saarbruecken voice database”. (2007), [Online]. Available: <http://stimddb.coli.uni-saarland.de>.
- [66] 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.
- [67] 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.

- 
- [68] P. Flandrin, “Time–frequency filtering based on spectrogram zeros”, *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2137–2141, 2015.
- [69] K. Gröchenig, *Foundations of time-frequency analysis*. Springer Science & Business Media, 2001.
- [70] 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.
- [71] P. Flandrin, “On spectrogram local maxima”, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2017, pp. 3979–3983.
- [72] D.-H. Pham and S. Meignen, “Demodulation algorithm based on higher order synchrosqueezing”, in *2019 27th European Signal Processing Conference (EUSIPCO)*, IEEE, 2019, pp. 1–5.
- [73] I. Daubechies, Y. G. Wang, and H.-T. Wu, “ConceFT: Concentration of frequency and time via a multitapered synchrosqueezed transform”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, 2016, ISSN: 1471-2962. DOI: [10.1098/rsta.2015.0193](https://doi.org/10.1098/rsta.2015.0193).
- [74] X.-G. Xia, “Discrete chirp-fourier transform and its application to chirp rate estimation”, *IEEE Transactions on Signal processing*, vol. 48, no. 11, pp. 3122–3133, 2000.
- [75] R. J. Baken and R. F. Orlikoff, *Clinical measurement of speech and voice*. Cengage Learning, 2000.
- [76] J. M. Miramont, M. A. Colominas, and G. Schlotthauer, “Voice jitter estimation using high-order synchrosqueezing operators”, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020.
- [77] S. Meignen and N. Singh, “Analysis of reassignment operators used in synchrosqueezing transforms: With an application to instantaneous frequency estimation”, *IEEE Transactions on Signal Processing*, vol. 70, pp. 216–227, 2021.
- [78] H. Holden, B. Øksendal, J. Ubøe, and T. Zhang, “Stochastic partial differential equations”, in *Stochastic partial differential equations*, Springer, 1996, pp. 141–191.
- [79] J. Pan and W. J. Tompkins, “A real-time qrs detection algorithm”, *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

- [80] D. T. Kaplan, “Simultaneous qrs detection and feature extraction using simple matched filter basis functions”, in *[1990] Proceedings Computers in Cardiology*, IEEE, 1990, pp. 503–506.
- [81] J. Huang, B. Chen, B. Yao, and W. He, “Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network”, *IEEE Access*, vol. 7, pp. 92 871–92 880, 2019.
- [82] B. S. Shaik, G. Naganjaneyulu, T Chandrasheker, and A. Narasimhadhan, “A method for qrs delineation based on stft using adaptive threshold”, *Procedia Computer Science*, vol. 54, pp. 646–653, 2015.
- [83] A. Quiceno-Manrique, J. Alonso-Hernandez, C. Travieso-Gonzalez, M. Ferrer-Ballester, and G Castellanos-Dominguez, “Detection of obstructive sleep apnea in ecg recordings using time-frequency distributions and dynamic features”, in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2009, pp. 5559–5562.
- [84] S. S. Qurraie and R. G. Afkhami, “Ecg arrhythmia classification using time frequency distribution techniques”, *Biomedical engineering letters*, vol. 7, no. 4, pp. 325–332, 2017.
- [85] M. Sharma and U. R. Acharya, “A new method to identify coronary artery disease with ecg signals and time-frequency concentrated antisymmetric biorthogonal wavelet filter bank”, *Pattern Recognition Letters*, vol. 125, pp. 235–240, 2019.
- [86] L. Su and H.-T. Wu, “Extract fetal ecg from single-lead abdominal ecg by de-shape short time fourier transform and nonlocal median”, *Frontiers in Applied Mathematics and Statistics*, vol. 3, p. 2, 2017.
- [87] N. Dia, J. Fontecave-Jallon, P.-Y. Gumery, and B. Rivet, “Fetal heart rate estimation from a single phonocardiogram signal using non-negative matrix factorization”, in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 5983–5986.
- [88] N. Dia *et al.*, “Fetal heart rate estimation by non-invasive single abdominal electrocardiography in real clinical conditions”, *Biomedical Signal Processing and Control*, vol. 71, p. 103 187, 2022.
- [89] C. Févotte and J. Idier, “Algorithms for nonnegative matrix factorization with the beta-divergence”, *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

- [90] C. Févotte, N. Bertin, and J.-L. Durrieu, “Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis”, *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.
- [91] V. Leplat, N. Gillis, and A. M. Ang, “Blind audio source separation with minimum-volume beta-divergence nmf”, *IEEE Transactions on Signal Processing*, vol. 68, pp. 3400–3410, 2020.
- [92] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, “A dynamical model for generating synthetic electrocardiogram signals”, *IEEE transactions on biomedical engineering*, vol. 50, no. 3, pp. 289–294, 2003.
- [93] S. Peleg and M. Werman, “Fast and robust earth mover’s distances”, in *IEEE Int. Conf. Computer. Vision.*, 2009, pp. 460–467.
- [94] A. Liutkus *et al.*, “The 2016 signal separation evaluation campaign”, in *International conference on latent variable analysis and signal separation*, Springer, 2017, pp. 323–332.
- [95] Y. Li and Q. He, “On the ratio of two correlated complex gaussian random variables”, *IEEE Communications Letters*, vol. 23, no. 12, pp. 2172–2176, 2019.



# Appendices



# A. Time Frequency Analysis of Interference

## A.1. Proof of Proposition 3.2.1

Let us consider the function  $(Ae^{-\pi\sigma^2(\eta-\xi_1)^2} + e^{-\pi\sigma^2(\eta-\xi_2)^2})^2$  having the same extrema as

$$l(\eta) = Ae^{-\pi\sigma^2(\eta-\xi_1)^2} + e^{-\pi\sigma^2(\eta-\xi_2)^2}.$$

Defining  $\xi = \frac{\xi_2 - \xi_1}{2}$  and making the change of variables  $\eta = \nu + \frac{\xi_1 + \xi_2}{2}$ , one gets:

$$l(\nu + \frac{\xi_1 + \xi_2}{2}) = Ae^{-\pi\sigma^2(\nu+\xi)^2} + e^{-\pi\sigma^2(\nu-\xi)^2},$$

which we denote by  $q(\nu)$ . Differentiating  $q$ , one obtains:

$$\begin{aligned} q'(\nu) &= -2\pi\sigma^2 e^{-\pi\sigma^2(\nu^2+\xi^2)} \\ &\quad \left[ A(\nu+\xi)e^{-2\pi\sigma^2\xi\nu} + (\nu-\xi)e^{2\pi\sigma^2\xi\nu} \right] \\ &= 2\pi\sigma^2 e^{-\pi\sigma^2\nu^2} e^{-\pi\sigma^2\xi^2} \xi (Ae^{-2\pi\sigma^2\xi\nu} + e^{2\pi\sigma^2\xi\nu}) \\ &\quad \left( -\frac{\nu}{\xi} - \frac{Ae^{-2\pi\sigma^2\xi\nu} - e^{2\pi\sigma^2\xi\nu}}{Ae^{-2\pi\sigma^2\xi\nu} + e^{2\pi\sigma^2\xi\nu}} \right) \\ &= 2\pi\sigma^2 e^{-\pi\sigma^2\nu^2} e^{-\pi\sigma^2\xi^2} \xi (Ae^{-2\pi\sigma^2\xi\nu} + e^{2\pi\sigma^2\xi\nu}) \\ &\quad \left( -\frac{\nu}{\xi} + \tanh(2\pi\sigma^2\xi\nu - \frac{\log(A)}{2}) \right). \end{aligned}$$

So  $q'(\nu)$  has the sign of  $h(\nu) := -\frac{\nu}{\xi} + \tanh(2\pi\sigma^2\xi\nu - \frac{\log(A)}{2})$  whose derivative is:

$$h'(\nu) = \frac{-1 + 2\pi\sigma^2\xi^2 \left( 1 - \tanh^2(2\pi\sigma^2\xi\nu - \frac{\log(A)}{2}) \right)}{\xi},$$



## APPENDIX A. TIME FREQUENCY ANALYSIS OF INTERFERENCE

---

which is negative if  $2\pi\sigma^2\xi^2 \leq 1$ . In such a case,  $q'$  is decreasing and then and as  $\lim_{\nu \rightarrow -\infty} q'(\nu) = +\infty$  and  $\lim_{\nu \rightarrow +\infty} q'(\nu) = -\infty$ , it annihilates once and thus  $l$  has a unique extremum which is a maximum.

Now when  $2\pi\sigma^2\xi^2 > 1$ ,  $h'$  has two zeros, and one has the following table of variations:

$\nu$	$-\infty$	$\nu_1$	$\nu_2$	$+\infty$
$h'(\nu)$	$-$	$0$	$+$	$0$
$h(\nu)$	$+\infty$	$h(\nu_1)$	$h(\nu_2)$	$-\infty$

If  $h(\nu_1) \geq 0$ , then  $h(\nu)$  and hence  $q'(\nu)$  annihilates once and changes signs at some  $\nu'_2$  in  $]\nu_2, +\infty[$ .  $q(\nu)$  is strictly increasing on  $]-\infty, \nu'_2]$  and strictly decreasing on  $[\nu'_2, +\infty[$ , and thus has a maximum at  $\nu = \nu'_2$  which is the unique extremum.

If  $h(\nu_2) \leq 0$ ,  $h(\nu)$  and hence  $q'(\nu)$  annihilates and changes signs once for a certain  $\nu'_1$  in  $]-\infty, \nu_1[$  and then  $q$  has a maximum at  $\nu = \nu'_1$  which is its unique extremum.

If  $h(\nu_1) < 0$  and  $h(\nu_2) > 0$ ,  $h(\nu)$  and  $q'(\nu)$  annihilate and change sign 3 times, once in  $]-\infty, \nu_1[$ , once in  $]\nu_1, \nu_2[$ , and once in  $[\nu_2, +\infty[$ , and thus  $q$  has three extrema.

To determine when  $q$  has three extrema we need to compute  $h(\nu_1)$  and  $h(\nu_2)$ . Remember that  $\nu_1$  and  $\nu_2$  are the roots of  $h'$  thus also of

$$1 - \frac{1}{2\pi\sigma^2\xi^2} = \tanh^2(2\pi\sigma^2\xi\nu - \frac{\log(A)}{2})$$

and therefore

$$\begin{aligned} \nu_1 &= \frac{1}{2\pi\sigma^2\xi} \left( \frac{\log(A)}{2} + \operatorname{atanh} \left( -\sqrt{1 - \frac{1}{2\pi\sigma^2\xi^2}} \right) \right) \\ \nu_2 &= \frac{1}{2\pi\sigma^2\xi} \left( \frac{\log(A)}{2} + \operatorname{atanh} \left( \sqrt{1 - \frac{1}{2\pi\sigma^2\xi^2}} \right) \right), \end{aligned}$$

and thus as  $\alpha = \sqrt{\frac{\pi}{2}}\sigma(\xi_2 - \xi_1) = \sqrt{2\pi}\sigma\xi$ , we may write:

$$\begin{aligned}
 h(\nu_1) &= -\frac{\log(A)}{4\pi\sigma^2\xi^2} - \frac{\operatorname{arctanh}(-\sqrt{1 - \frac{1}{2\pi\sigma^2\xi^2}})}{2\pi\sigma^2\xi^2} \\
 &\quad - \sqrt{1 - \frac{1}{2\pi\sigma^2\xi^2}} \\
 &= -\frac{\log(A)}{2\alpha^2} - \frac{\operatorname{arctanh}(-\frac{\sqrt{\alpha^2-1}}{\alpha})}{\alpha^2} - \frac{\sqrt{\alpha^2-1}}{\alpha} \\
 &= -\frac{\log(A)}{2\alpha^2} - \frac{\log(\frac{\alpha-\sqrt{\alpha^2-1}}{\alpha+\sqrt{\alpha^2-1}})}{2\alpha^2} - \frac{\sqrt{\alpha^2-1}}{\alpha} \\
 h(\nu_2) &= -\frac{\log(A)}{2\alpha^2} - \frac{\log(\frac{\alpha+\sqrt{\alpha^2-1}}{\alpha-\sqrt{\alpha^2-1}})}{2\alpha^2} + \frac{\sqrt{\alpha^2-1}}{\alpha}.
 \end{aligned}$$

From this we deduce that  $h(\nu_1) < 0$  and  $h(\nu_2) > 0$ , and thus  $q$  has three extrema, if and only if:

$$\begin{aligned}
 |\log(A)| &< 2\alpha\sqrt{\alpha^2-1} - \log\left(\frac{\alpha+\sqrt{\alpha^2-1}}{\alpha-\sqrt{\alpha^2-1}}\right) \\
 &= 2\alpha\sqrt{\alpha^2-1} - 2\operatorname{arccosh}(\alpha).
 \end{aligned}$$

## A.2. Proof of Proposition 3.2.2

To prove Proposition 3.2.2 one needs the following

**Lemma A.2.1.** *Assume  $\alpha$  is defined as in Proposition 3.2.1, and define  $\gamma = \cos(2\pi(\xi_2 - \xi_1)t)$ ,  $|V_f^g(t, \cdot)|^2$  has three extrema if and only if  $\alpha > \sqrt{\frac{1+\gamma}{2}}$  and*

$$|\log(A)| < -\operatorname{arccosh}(X_2) + 2\alpha^2 \frac{\sqrt{X_2^2-1}}{X_2+\gamma},$$

with  $X_2 = \gamma(\alpha^2 - 1) + \alpha\sqrt{\gamma^2(\alpha^2 - 2) + 2}$ . In any other case,  $|V_f^g(t, \cdot)|^2$  has a unique extremum.

*Proof.* Let us consider the function:

$$\begin{aligned}
 l(\eta) &= A^2 e^{-2\pi\sigma^2(\eta-\xi_1)^2} + e^{-2\pi\sigma^2(\eta-\xi_2)^2} \\
 &\quad + 2A e^{-\pi\sigma^2[(\eta-\xi_1)^2+(\eta-\xi_2)^2]} \cos(2\pi(\xi_2 - \xi_1)t),
 \end{aligned}$$

and then put  $\gamma = \cos(2\pi(\xi_2 - \xi_1)t)$ . Making the change of variables  $\eta = \nu + \frac{\xi_1+\xi_2}{2}$  and

putting  $\xi = \frac{\xi_2 - \xi_1}{2}$ , one may define:

$$\begin{aligned} f_1(\nu) &= l\left(\nu + \frac{\xi_1 + \xi_2}{2}\right) \\ &= e^{-2\pi\sigma^2(\nu^2 + \xi^2)} (A^2 e^{-4\pi\sigma^2\xi\nu} + e^{4\pi\sigma^2\xi\nu} + 2A\gamma). \end{aligned}$$

Putting  $\nu = \frac{\log(A) + \mu}{4\pi\sigma^2\xi}$ , enables us to define:

$$\begin{aligned} f_1(\nu) &= f_2(\mu) \\ &= e^{-2\pi\sigma^2 \left[ \left( \frac{\log(A) + \mu}{4\pi\sigma^2\xi} \right)^2 + \xi^2 \right]} (Ae^{-\mu} + Ae^{\mu} + 2A\gamma) \\ &= e^{-2\pi\sigma^2 \left[ \left( \frac{\log(A) + \mu}{4\pi\sigma^2\xi} \right)^2 + \xi^2 \right]} 2A(\cosh(\mu) + \gamma). \end{aligned}$$

The derivative of  $f_2$  reads:

$$\begin{aligned} f_2'(\mu) &= e^{-2\pi\sigma^2 \left[ \left( \frac{\log(A) + \mu}{4\pi\sigma^2\xi} \right)^2 + \xi^2 \right]} \\ &\quad 2A \left[ -\frac{\log(A) + \mu}{4\pi\sigma^2\xi^2} (\cosh(\mu) + \gamma) + \sinh(\mu) \right] \end{aligned}$$

which has the sign of (assuming  $\gamma \neq -1$ ):

$$\begin{aligned} h(\mu) &:= -\frac{\log(A) + \mu}{4\pi\sigma^2\xi^2} + \frac{\sinh(\mu)}{\cosh(\mu) + \gamma} \\ &= -\frac{\log(A) + \mu}{2\alpha^2} + \frac{\sinh(\mu)}{\cosh(\mu) + \gamma}. \end{aligned}$$

Differentiating  $h$  we get

$$\begin{aligned} h'(\mu) &= -\frac{1}{2\alpha^2} + \frac{\cosh(\mu)(\cosh(\mu) + \gamma) - \sinh^2(\mu)}{(\cosh(\mu) + \gamma)^2} \\ &= -\frac{1}{2\alpha^2} + \frac{1 + \gamma \cosh(\mu)}{(\cosh(\mu) + \gamma)^2} \end{aligned}$$

which has the same sign as

$$\begin{aligned} Y(\cosh(\mu)) &= -\frac{(\cosh(\mu) + \gamma)^2}{2\alpha^2} + 1 + \gamma \cosh(\mu) \\ &= -\frac{\cosh(\mu)^2}{2\alpha^2} + \gamma\left(1 - \frac{1}{\alpha^2}\right) \cosh(\mu) + 1 - \frac{\gamma^2}{2\alpha^2} \\ &= \frac{1}{\alpha^2} \left( -\frac{\cosh(\mu)^2}{2} + \gamma(\alpha^2 - 1) \cosh(\mu) + \alpha^2 - \frac{\gamma^2}{2} \right). \end{aligned}$$

The term inside the parentheses is a second order polynomial in  $\cosh(\mu)$  whose discriminant reads:  $\Delta = \alpha^2(\gamma^2(\alpha^2 - 2) + 2) > 0$ . The roots of this polynomial are denoted by  $X_1$  and  $X_2$  with  $X_1 < X_2$ , and  $Y(X) < 0$  if  $X \in ]-\infty, X_1[ \cup ]X_2, +\infty[$  and  $Y(X) > 0$  if  $X \in ]X_1, X_2[$ . To know the location of  $\cosh(\mu)$  with respect to  $X_1$  and  $X_2$ , one first compute  $Y(1) = (\gamma + 1)(1 - \frac{\gamma+1}{2\alpha^2})$  which has the same sign as  $\alpha^2 - \frac{\gamma+1}{2}$ . Assuming  $\alpha \leq \sqrt{\frac{\gamma+1}{2}}$ ,  $Y(1) \leq 0$  and 1 belongs to  $] -\infty, X_1[$  or  $]X_2, \infty[$ . As  $\frac{X_1+X_2}{2} = \gamma(\alpha^2 - 1) < 1$ , 1 belongs to  $]X_2, +\infty[$ . Finally, as  $\cosh(\mu) \geq 1$ ,  $Y(\cosh(\mu))$  and thus  $h'(\mu)$  are negative. From this study one deduces the following table of variations:

$\mu$	$-\infty$	$\mu_1$	$+\infty$
$h(\mu)$	$+\infty$		$-\infty$
$f'_2(\mu)$	+	0	-
$f_2(\mu)$			

and  $l(\eta)$  has a unique extremum which is a maximum when  $\alpha \leq \sqrt{\frac{\gamma+1}{2}}$ .

If one now assumes that  $\alpha > \sqrt{\frac{\gamma+1}{2}}$ , then  $Y(1) > 0$  and thus 1 belongs to  $]X_1, X_2[$ . As  $\cosh(\mu) \in ]X_1, X_2[$  is equivalent to  $|\mu| < \operatorname{arccosh}(X_2)$ ,  $Y(\cosh(\mu)) < 0$  if


$$\mu \in ]-\infty, -\operatorname{arccosh}(X_2)[ \cup ]\operatorname{arccosh}(X_2), +\infty[,$$

and  $Y(\cosh(\mu))$  is positive otherwise, leading to the following table of variations:

$\mu$	$-\infty$	$-\operatorname{arccosh}(X_2)$	$\operatorname{arccosh}(X_2)$	$+\infty$
$h'(\mu)$	-	0	+	0
$h(\mu)$	$+\infty$			$-\infty$

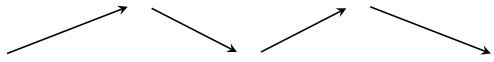
$h(\mu)$  vanishes with a change of sign only once at some  $\mu = \mu_1$  if and only if  $h(-\operatorname{arccosh}(X_2)) \geq$

0 or  $h(\operatorname{arccosh}(X_2)) \leq 0$ . In this case, we deduce that

$\mu$	$-\infty$	$\mu_1$	$+\infty$	
$f_2'(\mu)$		+	0	-
$f_2(\mu)$				

meaning  $l(\eta)$  has a unique extremum (which is a maximum).

If  $h(-\operatorname{arccosh}(X_2)) < 0$  and  $h(\operatorname{arccosh}(X_2)) > 0$ ,  $h(\mu)$  vanishes and changes signs three times at some  $\mu = \mu_1, \mu_2$  and  $\mu_3$  leading to the following table of variations for  $f_2$ :

$\mu$	$-\infty$	$\mu_1$	$\mu_2$	$\mu_3$	$+\infty$			
$h(\mu)$		+	0	-	0	+	0	-
$f_2(\mu)$								

In such case,  $l(\eta)$  has three extrema: 2 maxima and a minimum.

Finally, to specify in which situations  $l$  has three extrema, we remark that

$$\begin{aligned}
 & h(\operatorname{arccosh}(X_2)) > 0 \\
 \Leftrightarrow & \log(A) < -\operatorname{arccosh}(X_2) + 2\alpha^2 \frac{\sqrt{X_2^2 - 1}}{X_2 + \gamma} \\
 & h(-\operatorname{arccosh}(X_2)) < 0 \\
 \Leftrightarrow & -\log(A) < -\operatorname{arccosh}(X_2) + 2\alpha^2 \frac{\sqrt{X_2^2 - 1}}{X_2 + \gamma}.
 \end{aligned}$$

So  $l(\eta)$  has three extrema if and only if  $\alpha > \sqrt{\frac{1+\gamma}{2}}$  and then if  $|\log(A)| < -\operatorname{arccosh}(X_2) + 2\alpha^2 \frac{\sqrt{X_2^2 - 1}}{X_2 + \alpha}$  with  $X_2 = \gamma(\alpha^2 - 1) + \alpha\sqrt{\gamma^2(\alpha^2 - 2) + 2}$ .  $\square$

To prove Proposition 3.2.2, assume that  $-1 < \gamma \leq 1$ .  $f_2'(\mu)$  has the sign of:

$$h(\mu) = -\frac{\log(A) + \mu}{4\pi\sigma^2\xi^2} + \frac{\sinh(\mu)}{\cosh(\mu) + \gamma},$$

which one denote by  $h(\mu, \gamma)$ , rewritten as:

$$h(\mu, \gamma) = -\frac{\log(A) + \mu}{2\alpha^2} + \frac{\sinh(\mu)}{\cosh(\mu) + \gamma}.$$

We know from the lemma that  $f_2$  has three extrema if and only if  $\alpha > \sqrt{\frac{1+\gamma}{2}}$  and  $h(-\operatorname{arccosh}(X_2), \gamma) < 0$  and  $h(\operatorname{arccosh}(X_2), \gamma) > 0$ . Let us show that this is equivalent

to:

$$\exists y_0 < 0 \ h(y_0, \gamma) < 0 \text{ and } y'_0 > 0 \ h(y'_0, \gamma) > 0.$$

This condition is a necessary condition since the values  $y_0 = -\operatorname{arccosh}(X_2)$  and  $y'_0 = \operatorname{arccosh}(X_2)$ , satisfy the property. Conversely, let us assume that there exist  $y_0 < 0$  and  $y'_0 > 0$  such that  $h(y_0, \gamma) < 0$  and  $h(y'_0, \gamma) > 0$ . As  $\lim_{\mu \rightarrow +\infty} h(\mu, \gamma) = -\infty$  and  $\lim_{\mu \rightarrow -\infty} h(\mu, \gamma) = +\infty$ ,  $h(\cdot, \gamma)$  (and also  $f'_2(\mu)$ ) vanishes and changes signs three times, respectively on the intervals  $] -\infty, y_0[$ ,  $]y_0, y'_0[$ , and on  $]y'_0, +\infty[$ , and thus  $f_2$  has three extrema.

Let us now show that, if the function  $f_2$  has three extrema for  $\gamma = 1$ , it also has three extrema for  $\gamma \in ]-1, 1]$ . Based on the characterization using the function  $h$  of the existence of three extrema, assume there exist  $y_0 < 0$  and  $y'_0 > 0$  such that  $h(y_0, 1) < 0$  and  $h(y'_0, 1) > 0$ . Then as  $\gamma \in ]-1, 1]$  one has  $\forall \mu \in \mathbb{R} \ \frac{1}{\cosh(\mu)} \geq \frac{1}{\cosh(\mu)+1}$ , which means that  $\forall \mu \geq 0$ ,  $\frac{\sinh(\mu)}{\cosh(\mu)+\gamma} \geq \frac{\sinh(\mu)}{\cosh(\mu)+1}$  and  $\forall \mu \leq 0$ ,  $\frac{\sinh(\mu)}{\cosh(\mu)+\gamma} \leq \frac{\sinh(\mu)}{\cosh(\mu)+1}$ . Finally, one deduces that  $\forall \mu \geq 0$ ,  $h(\mu, \gamma) \geq h(\mu, 1)$  and  $\forall \mu \leq 0$ ,  $h(\mu, \gamma) \leq h(\mu, 1)$ , which means in particular that  $h(y'_0, \gamma) \geq h(y'_0, 1)$ , and that  $h(y_0, \gamma) \leq h(y_0, 1)$ , implying that  $h(y'_0, \gamma) > 0$  and  $h(y_0, \gamma) < 0$ . Thus whatever  $\gamma \in ]-1, 1]$ , the function  $h$  satisfies the condition for  $f_2$  to have three extrema.

So finally, we may conclude that the function  $f_2$  has three extrema whatever  $\gamma \in [-1, 1]$ , if and only if  $\alpha > 1$  and  $|\log(A)| < -2 \operatorname{arccosh}(\alpha) + 2\alpha\sqrt{\alpha^2 - 1}$ .



## B. Towards Robust and Adaptive Synchrosqueezing

### B.1. Proof of Proposition 5.0.1

In what follows, we omit  $(t, \eta)$  in the definition of the **STFTs** (STFTs!). Let us first remark that:

$$\mathbb{E}\{V_{f+\varepsilon}^g\} = V_f^g, \quad \mathbb{E}\{V_{f+\varepsilon}^{t^2g}\} = V_f^{t^2g}$$

since  $V_\varepsilon^g$  and  $V_\varepsilon^{t^2g}$  are both with zero mean. Let us denote by  $\mathbf{m} = (V_f^g, V_f^{t^2g})^T$ . Then denoting by  $\mathbf{v} = (V_{f+\varepsilon}^g, V_{f+\varepsilon}^{t^2g})^T$ , and  $\sigma_\varepsilon^2$  the variance of the noise (the real and imaginary part of the noise are assumed to be decorrelated), we have:

$$\mathbb{E}\{(\mathbf{v} - \mathbf{m})(\mathbf{v} - \mathbf{m})^H\} = \begin{bmatrix} \mathbb{E}\{|V_\varepsilon^g|^2\} & \mathbb{E}\{V_\varepsilon^g(V_\varepsilon^{t^2g})^*\} \\ \mathbb{E}\{(V_\varepsilon^g)^*V_\varepsilon^{t^2g}\} & \mathbb{E}\{|V_\varepsilon^{t^2g}|^2\} \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2\|g\|^2 & \sigma_\varepsilon^2\|tg\|^2 \\ \sigma_\varepsilon^2\|tg\|^2 & \sigma_\varepsilon^2\|t^2g\|^2 \end{bmatrix}.$$

Based on the results in [95, Eq. (44)], denoting  $Z = \frac{V_{f+\varepsilon}^g}{V_{f+\varepsilon}^{t^2g}}$ , we have that:

$$\mathbb{E}\{Z\} = \frac{V_f^g}{V_f^{t^2g}} + \left( \frac{\|tg\|^2}{\|t^2g\|^2} - \frac{V_f^g}{V_f^{t^2g}} \right) e^{-\frac{|V_f^{t^2g}|^2}{\sigma_\varepsilon^2\|t^2g\|^2}},$$

so that one gets:

$$\mathbb{E}\left\{-\frac{1}{2\pi}\Im\{Z\}\right\} = -\frac{1}{2\pi}\Im\left\{\frac{V_f^g}{V_f^{t^2g}}\right\} \left(1 - e^{-\frac{|V_f^{t^2g}|^2}{\sigma_\varepsilon^2\|t^2g\|^2}}\right).$$



At  $(t, \phi'(t))$  one finally has

$$\mathbb{E} \left\{ -\frac{1}{2\pi} \Im \{Z\} \right\} = \phi''(t) \left( 1 - e^{-\frac{|V_f^{t^2} g|^2}{\sigma_\varepsilon^2 \|t^2 g\|^2}} \right).$$

## B.2. Proof of Proposition 5.0.2

Remarking that in that case  $\phi'(t) = \eta_0$ , we may write:

$$\begin{aligned} V_\varepsilon^{t^2 g}(t, \eta_0) e^{-2i\pi\eta_0 t} &= \int_{\mathbb{R}} n(\tau) (\tau - t)^2 g(\tau - t) e^{-2i\pi\eta_0 \tau} d\tau \\ V_\varepsilon^g(t, \eta_0) e^{-2i\pi\eta_0 t} &= \int_{\mathbb{R}} n(\tau) g(\tau - t) e^{-2i\pi\eta_0 \tau} d\tau, \end{aligned}$$

so we get that:

$$\begin{aligned} G(t) &= \frac{2\pi}{\sigma^2} V_\varepsilon^{t^2 g}(t, \eta_0) e^{-2i\pi\eta_0 t} - V_\varepsilon^g(t, \eta_0) e^{-2i\pi\eta_0 t} \\ &= \int_{\mathbb{R}} n(\tau) \left( \frac{2\pi}{\sigma^2} (\tau - t)^2 - 1 \right) g(\tau - t) e^{-2i\pi\eta_0 \tau} d\tau \\ &= \frac{\sigma^2}{2\pi} \int_{\mathbb{R}} n(\tau) g''(\tau - t) e^{-2i\pi\eta_0 \tau} d\tau. \end{aligned}$$

Let us then write the auto-correlation function of this random variable:

$$\begin{aligned} \mathbb{E} [G(t) G(t - x)^*] &= \int_{\mathbb{R}^2} \mathbb{E} [n(\tau) n(\tau')] g''(\tau) g''(\tau' + x) \frac{\sigma^4 e^{-2i\pi\eta_0(\tau - \tau')}}{4\pi^2} d\tau d\tau' \\ &= \frac{\sigma^4}{4\pi^2} \sigma_\varepsilon^2 \int_{\mathbb{R}} g''(\tau) g''(\tau + x) d\tau \\ &= \frac{\sigma^4}{4\pi^2} \sigma_\varepsilon^2 g'' * g''(-x), \end{aligned}$$

using the fact that  $g''$  is even. So the studied process is wide-sense stationary and its power spectral density, i.e. the Fourier transform of the auto-correlation function, reads:

$$\begin{aligned} \mathcal{F}_{\frac{\sigma^4}{4\pi^2} \sigma_\varepsilon^2 g'' * g''(-x)}(\eta) &= \sigma_\varepsilon^2 \frac{\sigma^4}{4\pi^2} \mathcal{F}_{g'' * g''}(\eta)^* \\ &= \sigma_\varepsilon^2 \frac{\sigma^4}{4\pi^2} \widehat{g''}(\eta)^2 \\ &= \sigma_\varepsilon^2 \sigma^4 4\pi^2 \eta^4 \widehat{g}(\eta)^2 \\ &= \sigma_\varepsilon^2 \sigma^6 4\pi^2 \eta^4 e^{-2\pi\sigma^2 \eta^2}. \end{aligned}$$

### B.3. Proof of Proposition 5.0.3

In such a case, we have  $\phi'(t) = a + bt$ , and thus we may write:

$$\begin{aligned} V_\varepsilon^{t^2} g(t, \phi'(t)) e^{-2i\pi\phi(t)} &= e^{i\pi bt^2} \int_{\mathbb{R}} \varepsilon(\tau) (\tau - t)^2 g(\tau - t) e^{-2i\pi\phi'(t)\tau} d\tau \\ V_\varepsilon^g(t, \phi'(t)) e^{-2i\pi\phi(t)} &= e^{i\pi bt^2} \int_{\mathbb{R}} \varepsilon(\tau) g(\tau - t) e^{-2i\pi\phi'(t)\tau} d\tau, \end{aligned}$$

Furthermore, from [19, proposition 1], we may write:

$$V_f^g(t, \phi'(t)) = (1 + b^2\sigma^4)^{-\frac{1}{4}} e^{-i\frac{atan(-b\sigma^2)}{2}} e^{2i\pi\phi(t)} = B e^{2i\pi\phi(t)},$$

so that one may rewrite  $G$  as:

$$\begin{aligned} G(t) &= \left( \frac{2\pi}{\sigma^2} (1 - ib\sigma^2) V_\varepsilon^{t^2} g - V_\varepsilon^g \right) B^{-1} e^{-2i\pi\phi(t)} \\ &= B^{-1} e^{i\pi bt^2} \int_{\mathbb{R}} \varepsilon(\tau) \left( \frac{2\pi}{\sigma^2} (1 - ib\sigma^2) (\tau - t)^2 - 1 \right) g(\tau - t) e^{-2i\pi\phi'(t)\tau} d\tau \\ &= B^{-1} e^{i\pi bt^2} \left[ \frac{\sigma^2}{2\pi} \int_{\mathbb{R}} \varepsilon(\tau) g''(\tau - t) e^{-2i\pi\phi'(t)\tau} d\tau - i2\pi b \int_{\mathbb{R}} \varepsilon(\tau) (\tau - t)^2 g(\tau - t) e^{-2i\pi\phi'(t)\tau} d\tau \right]. \end{aligned}$$

Then the auto-correlation function of  $G$  reads:

$$\begin{aligned} \mathbb{E}[G(t)G(t-x)^*] &= \frac{e^{-i\pi bx^2} \sigma_\varepsilon^2}{(1 + b^2\sigma^4)^{\frac{1}{2}}} \\ &\quad \left\{ \frac{\sigma^4}{4\pi^2} \int_{\mathbb{R}} g''(\tau) g''(\tau + x) e^{-2i\pi bx\tau} d\tau \right. \\ &\quad \quad ib\sigma^2 \int_{\mathbb{R}} \tau^2 g(\tau) g''(\tau + x) e^{-2i\pi bx\tau} d\tau \\ &\quad \quad - ib\sigma^2 \int_{\mathbb{R}} (\tau + x)^2 g''(\tau) g(\tau + x) e^{-2i\pi bx\tau} d\tau \\ &\quad \quad \left. 4\pi^2 b^2 \int_{\mathbb{R}} \tau^2 g(\tau) (\tau + x)^2 g(\tau + x) e^{-2i\pi bx\tau} d\tau \right\}. \end{aligned}$$

So the process is wide-sense stationary. Now, making a last change of variables, we obtain:

$$\begin{aligned} \mathbb{E}[G(t)G(t-x)^*] &= \frac{\sigma_\varepsilon^2}{(1+b^2\sigma^4)^{\frac{1}{2}}} \\ &\quad \left\{ \frac{\sigma^4}{4\pi^2} \int_{\mathbb{R}} g''(\tau - \frac{x}{2})g''(\tau + \frac{x}{2})e^{-2i\pi b x \tau} d\tau \right. \\ &\quad + ib\sigma^2 \int_{\mathbb{R}} (\tau - \frac{x}{2})^2 g(\tau - \frac{x}{2})g''(\tau + \frac{x}{2})e^{-2i\pi b x \tau} d\tau \\ &\quad - ib\sigma^2 \int_{\mathbb{R}} (\tau + \frac{x}{2})^2 g''(\tau - \frac{x}{2})g(\tau + \frac{x}{2})e^{-2i\pi b x \tau} d\tau \\ &\quad \left. + 4\pi^2 b^2 \int_{\mathbb{R}} (\tau - \frac{x}{2})^2 g(\tau - \frac{x}{2})(\tau + \frac{x}{2})^2 g(\tau + \frac{x}{2})e^{-2i\pi b x \tau} d\tau \right\}. \end{aligned}$$

Now, one can check that the auto-correlation function can be rewritten in terms of the function:

$$L(\tau) = \left( \frac{\sigma^2}{2\pi} g''(\tau) - i2\pi b \tau^2 g(\tau) \right) e^{i\pi b \tau^2}$$

since one has:

$$\mathbb{E}[G(t)G(t-x)^*] = \frac{\sigma_\varepsilon^2}{(1+b^2\sigma^4)^{\frac{1}{2}}} \int_{\mathbb{R}} L(\tau - \frac{x}{2})L(\tau + \frac{x}{2})^* d\tau.$$

So, computing the Fourier transform of this auto-correlation, one gets:

$$\begin{aligned} &\int_{\mathbb{R}} \mathbb{E}[G(t)G(t-x)^*] e^{-2i\pi x \eta} dx \\ &= \frac{\sigma_\varepsilon^2}{(1+b^2\sigma^4)^{\frac{1}{2}}} \int_{\mathbb{R}} \int_{\mathbb{R}} L(\tau - \frac{x}{2})L(\tau + \frac{x}{2})^* e^{-2i\pi x \eta} dx d\tau \\ &= \frac{\sigma_\varepsilon^2}{(1+b^2\sigma^4)^{\frac{1}{2}}} \int_{\mathbb{R}} W_L(\tau, \eta) d\tau, \end{aligned}$$

in which  $W_L$  is the Wigner-Ville distribution of  $L$ . From the properties of the Wigner-Ville distribution, we deduce that the modulus squared of the Fourier transform of  $L$  reads:

$$|\widehat{L}(\eta)|^2 = \int_{\mathbb{R}} W_L(\tau, \eta) d\tau.$$

Now as we use a Gaussian window we can have an analytic expression for  $\widehat{L}$ . Indeed, some simple computations lead to:

$$\begin{aligned} L(\tau) &= \left( \left( \frac{2\pi}{\sigma^2} \tau^2 - 1 \right) - i2\pi b \tau^2 \right) g(\tau) e^{i\pi b \tau^2} \\ &= \left( -2\pi \left( -\frac{1}{\sigma^2} + ib \right) \tau^2 - 1 \right) e^{\pi \left( -\frac{1}{\sigma^2} + ib \right) \tau^2} \end{aligned}$$

So, if one considers the Fourier transform of this function, one gets:

$$\begin{aligned} \widehat{L}(\eta) &= \frac{-1+ib\sigma^2}{2\pi\sigma^2} \frac{d^2 \mathcal{F}_{e^{-\pi(\frac{1-ib\sigma^2}{\sigma^2})\tau^2}}}{d\eta^2}(\eta) - \mathcal{F}_{e^{-\pi(\frac{1-ib\sigma^2}{\sigma^2})\tau^2}}(\eta) \\ &= -\frac{(1-ib\sigma^2)^{\frac{1}{2}}}{2\pi\sigma} \frac{d^2 e^{-\pi(\frac{\sigma^2}{1-ib\sigma^2})\eta^2}}{d\eta^2} - \frac{\sigma e^{-\pi(\frac{\sigma^2}{1-ib\sigma^2})\eta^2}}{(1-ib\sigma^2)^{\frac{1}{2}}} \\ &= -\frac{\sigma^3 2\pi}{(1-ib\sigma^2)^{\frac{3}{2}}} \eta^2 e^{-\pi(\frac{\sigma^2}{1-ib\sigma^2})\eta^2}. \end{aligned}$$

From this, we finally deduce that the power spectral density of  $G_1$  is:

$$P_{G_1}(\eta) = \frac{\sigma_\varepsilon^2 \sigma^6 4\pi^2 \eta^4}{(1+b^2\sigma^4)^2} e^{-\frac{2\pi\sigma^2\eta^2}{1+b^2\sigma^4}}.$$



## Abstract

Time-frequency analysis is commonly used to study real world signals. These can often be described as multicomponent signals made of the sum of frequency and amplitude modulated modes. This thesis describes time-frequency techniques to study such signals in heavy noise situations, by dealing with three major issues. The first one is to design robust ridge detection technique and linear chirp approximation to improve instantaneous frequency estimation and mode reconstruction. The results demonstrate that an accurate ridge detection is a necessary but not a sufficient condition to ensure an accurate mode reconstruction. The second issue is the identification and the separation of interfering modes. To address this issue, the approach we propose focuses on ridge detection to localize patterns, coined time frequency bubbles, associated with interference in the time-frequency plane. The third issue is on the adaption of the synchrosqueezing transform to the frequency modulation of the modes and to noise. Regarding the first aspect, an energy based criteria is defined to measure the concentration of a time-frequency representation, which we use to adapt the synchrosqueezing technique. On the noise issue, based on a theoretical study of the effect of noise on the chirp rate estimator used in synchrosqueezing transforms, a new chirp rate denoising technique is proposed improving the estimation. The fourth and last issue is the heart rate estimation on ECG signals using time-frequency analysis, for which we design a specific algorithm. We show that the choice of the representation has huge consequences and demonstrate that it should be taken into account.

## Résumé

L'analyse temps-fréquence est fréquemment utilisée pour étudier les signaux du monde réel. Ceux-ci peuvent souvent être décrits comme des signaux multicomposantes, représentés sous la forme d'une somme de modes, modulés en amplitude et en fréquence. Cette thèse traite de techniques temps-fréquence pour l'étude de tels signaux dans des situations où le niveau de bruit est élevé pour trois principales problématiques. La première est l'élaboration de détection de ligne de crêtes et l'approximation basées sur un modèle de chirp linéaire dans l'objectif de gagner en précision sur l'estimation de la fréquence instantanée et la reconstruction des modes. La deuxième problématique est l'identification et la séparation de modes qui interfèrent. Cette problématique est adressée, nous proposons une approche basée sur la détection de ligne de crête pour identifier des structures appelées bulles temps-fréquence, associées à l'interférence sur le plan temps-fréquence. La troisième problématique porte sur l'adaptation du synchrosqueezing à la modulation en fréquence des modes ainsi qu'au bruit. Sur ce premier aspect, nous définissons un critère basé sur l'énergie pour mesurer la concentration des représentations temps-fréquence, que nous utilisons pour adapter le synchrosqueezing. Concernant le problème du bruit, en nous basant sur l'étude théorique de l'effet du bruit sur l'estimateur de la modulation en fréquence, utilisée pour le synchrosqueezing, nous proposons une nouvelle technique de débruitage de manière à obtenir une meilleure estimation. Nous traitons une quatrième problématique, qui concerne l'estimation de la fréquence cardiaque en utilisant une approche temps-fréquence, pour laquelle nous construisons un algorithme. Nous montrons que le choix de la représentation a de grandes conséquences et qu'elle doit être prise en compte.