HAL
open science

Retrieval of the Modes of Multicomponent Signals from Downsampled Short-Time Fourier Transform

Sylvain Meignen, Duong-Hung Pham

To cite this version:

Sylvain Meignen, Duong-Hung Pham. Retrieval of the Modes of Multicomponent Signals from Downsampled Short-Time Fourier Transform. IEEE Transactions on Signal Processing, 2018, 66 (23), pp.6204-6215. 10.1109/TSP.2018.2875390 . hal-01850940

HAL Id: hal-01850940

https://hal.science/hal-01850940

Submitted on 28 Jul 2018

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.

On Multicomponent Signal Denoising and Mode Reconstruction from Short-Time Fourier Transform

Sylvain Meignen, and Duong-Hung Pham

Abstract

In this paper, we investigate the reconstruction and denoising of multicomponent signals from their downsampled short-time Fourier transform. We first recall signal reconstruction techniques based on short-time Fourier transforms and how to use them for signal denoising. We then investigate how to adapt the proposed denoising technique to the context of multicomponent signals. Comparisons with recently developed techniques based on synhrosqueezing conclude the paper.

Index Terms

Time-frequency, signal denoising, short-time Fourier transform, downsampling, reassignment, AM/FM multicomponent signals.

I. Introduction

Time-frequency (TF) representations such as the short-time Fourier transform (STFT) or the continuous wavelet transform (CWT) are widely used to represent multi-component signals (MCSs). Doing some kind of filtering, they smear the information contained in the signal. A variety of techniques have been developed to recover the TF signature associated with a MCS, as for instance *reassignment methods* [1] of which synchrosqueezing transforms (SST) are among the most popular [2]. Indeed, not only do these transforms enhance the TF representation but also allow for mode reconstruction. Originally designed to reassign the TF representations of MCSs containing slightly modulated modes, either in the CWT

S. Meignen and D-H Pham are with the Jean Kuntzmann Laboratory, University of Grenoble-Alpes, and CNRS, Grenoble 38041, France (email:duong-hung.pham@imag.fr and sylvain.meignen@imag.fr).

or STFT contexts [2] [3] [4], several extensions were recently proposed to better take into account the modulation either by demodulating the signal [5], by taking into account phase and amplitude modulations in the synchrosqueezing operators [6] [7] or both [8].

Despite some theoretical analyses have been carried out [9], the impact of noise on synchrosqueezing operators still needs to be further studied, in particular in heavy noise situations. From a practical point of view, different techniques have been developed to handle the noise in the SST context, among which a very promising one, inspired by multitaper approaches [10], is based on averaging the SSTs obtained with different kernels [11]. Another issue related to the use of SST is that while the latter is sharper than the TF representation it is based on, mode reconstruction cannot be carried out from SST downsampled in time. On the contrary, such a reconstruction can be achieved from a downsampled STFT: in some sense, SST, by sharpening the frequency representation, annihilates the time redundancy of the TF representation it is based on. Our goal is thus to show that, using the time redundancy of STFT, the reconstruction of the modes of a noisy MCS can be performed with a better accuracy and using less STFT coefficients than when STFT based SST is considered.

For that purpose, in Section II, we recall the different signal reconstruction techniques from shifted downsampled STFT, and then study how to use these for signal denoising in Section III, assuming the noise is Gaussian white. Exploiting these findings, we introduce, in Section IV, a novel algorithm, based on downsampled STFT, for the retrieval of the modes of a MCS, of which we study, in the next section, the sensitivity to its parameters. We finally carry out a comparison with two other methods based on SST and recently introduced in [6] [8], on both simulated and real data.

II. Signal Reconstruction from Downsampled Short-Time Fourier Transform

For a signal f in $l_2(\mathbb{Z})$, and a discrete window g also in $l_2(\mathbb{Z})$, the STFT downsampled by a factor of R (a positive integer) and shifted by an integer parameter p ($0 \leq p < R$) is defined for each ω by:

$$V_f^g(mR+p,\omega) \quad = \quad \sum_{n\in\mathbb{Z}} f[n]g[n-mR-p]e^{-i2\pi\omega(n-mR-p)}. \tag{1}$$

If one assumes the signal is of length L and supported on $\{0,\cdots,L-1\}$, the filter g supported on $\{-M,\cdots,M\}$ such that $L_g := 2M+1 \leq N \leq L/2$, where N is the number of frequency bins, one

may write:

$$\begin{aligned} V_f^g(mR+p,\frac{k}{N}) &= \sum_{n\in\mathbb{Z}} f[n]g[n-mR-p]e^{-i2\pi\frac{k(n-mR-p)}{N}} \\ &= \sum_{n=-M}^{M} f[mR+p+n]g[n]e^{-i2\pi\frac{kn}{N}} \\ &= \sum_{n=0}^{2M} f[mR+p+n-M]g[n-M]e^{-i2\pi\frac{k(n-M)}{N}}. \end{aligned}$$

Since g is null on $\{M+1,\cdots,N-1-M\}$, STFT can be rewritten as:

$$V_f^g(mR+p,\frac{k}{N})e^{-i2\pi\frac{kM}{N}} = \sum_{n=0}^{N-1} f[mR+p+n-M]g[n-M]e^{-i2\pi\frac{kn}{N}}.$$

Using the properties of the *discrete Fourier transform* (DFT), one obtains, for any $n\in\{0,\cdots,N-1\}$:

$$f[mR+p+n-M]g[n-M] = \frac{1}{N}\sum_{k=0}^{N-1} V_f^g(mR+p,\frac{k}{N})e^{i2\pi\frac{k(n-M)}{N}}. \tag{2}$$

When g is non-zero on $\{-p,\cdots,R-1-p\}$, one gets, for any $q\in\{0,\cdots,R-1\}$:

$$f[mR+q] = \frac{1}{g[q-p]N}\sum_{k=0}^{N-1} V_f^g(mR+p,\frac{k}{N})e^{i2\pi\frac{k(q-p)}{N}}. \tag{3}$$

This reconstruction technique will be referred to as the *first reconstruction technique* (RT1) in the sequel. Since p varies in $\{0,\cdots,R-1\}$, if one wants to use the same filter g whatever p, the former has to be non-zero on $\{-R+1,\cdots,R-1\}$.

Alternatively, using (2) and assuming $\sum_{m\in\mathbb{Z}} g[n-mR]^2 = 1$ for all n, we may write:

$$f[n] = \sum_{k=0}^{N-1}\sum_{m\in\mathbb{Z}} V_f^g(mR+p,\frac{k}{N})g[n-mR-p]\frac{e^{i2\pi\frac{k(n-mR-P)}{N}}}{N}. \tag{4}$$

The proof of (4) is available in Section Appendix. Furthermore, when such a filter is used, one also has energy conservation, namely:

$$\sum_{n\in\mathbb{Z}} f[n]^2 = \sum_{m\in\mathbb{Z}}\frac{1}{N}\sum_{k=0}^{N-1}|V_f^g(mR+p,\frac{k}{N})|^2, \tag{5}$$

the proof being also available in Section Appendix. Note that the reconstruction can still be carried out by making the appropriate renormalization as follows:

$$f[n] = \frac{\sum_{k=0}^{N-1}\sum_{m\in\mathbb{Z}} V_f^g(mR+p,\frac{k}{N})g[n-mR-p]\frac{e^{i2\pi\frac{k(n-mR-p)}{N}}}{N}}{\sum_{m\in\mathbb{Z}} g[n-mR-p]^2}, \tag{6}$$

provided $\sum_{m\in\mathbb{Z}} g[n-mR]^2 \neq 0$ for all n. This type of reconstruction technique is denoted *second reconstruction technique* (RT2) in the sequel.

Remark 1. *It is important to remark that such a reconstruction formula was already studied in [12] but for infinite signals and assuming a continuous frequency. Here, the signal is finite and the frequencies are discretized. Also, in our context the number of frequency bins is much smaller than the signal length for the method to be tractable for long signals.*

Alternatively, assuming this time $\sum_{m\in\mathbb{Z}} g[n-mR]=1$ for all n, one similarly has:

$$f[n]=\sum_{k=0}^{N-1}\sum_{m\in\mathbb{Z}} V_f^g(mR+p,\frac{k}{N})\frac{e^{i2\pi\frac{k(n-mR-p)}{N}}}{N}, \tag{7}$$

the proof being available in Section Appendix. Note that, as soon as $\sum_{m\in\mathbb{Z}} g[n-mR]\neq 0$ for all n, one can still reconstruct f through:

$$f[n]=\frac{\sum_{k=0}^{N-1}\sum_{m\in\mathbb{Z}} V_f^g(mR+p,\frac{k}{N})\frac{e^{i2\pi\frac{k(n-mR-p)}{N}}}{N}}{\sum_{m\in\mathbb{Z}} g[n-mR-p]}. \tag{8}$$

This reconstruction formula is referred to as *third reconstruction technique* (RT3) in the sequel.

Remark 2. *Typical filters g that can be used in RT2 and RT3 are those satisfying the* constant overlap add *property [13] with hop-size R ($COLA(R)$): $\sum_{m\in\mathbb{Z}} g[n-mR]=1$, for all n.*

For example, the rectangular window defined on $\{0,\cdots,R-1\}$ is $COLA(R)$. Other popular examples of $COLA(R)$ windows is the Bartlett window: $w[n]=w_r[n]\left[1-\frac{|n|}{R}\right]$, for $n\in\{-R,\cdots,R\}$ and where w_r is the rectangular window on $\{-R,\cdots,R\}$, the Hann window defined, for $n\in\{-R,\cdots,R\}$, by: $w[n]=w_r[n]\left[\frac{1}{2}+\frac{1}{2}\cos(\frac{\pi n}{R})\right]$, or the Hamming window, $w[n]=w_r[n]\left[0.54+(1-0.54)\cos(\frac{\pi n}{R})\right]$, for this time $n\in\{-R,\cdots,R-1\}$, and putting $w[R]=0$. One has however to bear in mind that $COLA(R)$ condition is only sufficient for signal reconstruction from downsampled STFT.

III. Signal Denoising from Downsampled STFT

Let us consider a discrete-time noisy signal $\tilde{f}=f+\sigma\Phi$ where Φ is a unit variance, zero mean white Gaussian noise. When the signal samples are complex numbers, the noise is also complex, and Φ is such that $V_\Phi^g(m,\frac{k}{N})$ is also Gaussian with zero mean and satisfies:

$$\text{Var}\left(\Re\left\{V_\Phi^g(m,\frac{k}{N})\right\}\right)=\text{Var}\left(\Im\left\{V_\Phi^g(m,\frac{k}{N})\right\}\right)=\sigma^2\|g\|_2^2, \tag{9}$$

where $\Re\{X\}$ (resp. $\Im\{X\}$) denotes the real (resp. imaginary) part of complex number X, and Var the variance. From this, we deduce that $\left(|V_\Phi^g(m,\frac{k}{N})|^2)_{k,m}\right)$ is χ_2 distributed with 2 degrees of freedom. Because of (9), the expectation of $|V_\Phi^g(m,\frac{k}{N})|^2$ is $2\sigma^2\|g\|_2^2$. Now, the probability of false alarm is less

than 1% if $|V_{\Phi}^{g}(m, \frac{k}{N})|^2$ is larger than $9 \times 2\sigma^2 \|g\|_2^2$. Therefore, assuming the variance of the noise σ^2 is known, we define the following thresholding strategy:

$$\overline{V}_{\tilde{f}}^{g}(m, \frac{k}{N}) = \begin{cases} V_{\tilde{f}}^{g}(m, \frac{k}{N}), \text{if } |V_{\tilde{f}}^{g}(m, \frac{k}{N})| \geq 3\sqrt{2}\sigma\|g\|_2 \\ 0 \text{ otherwise,} \end{cases} \tag{10}$$

where $\|g\|_2$ is the l_2 norm of filter g. Denoised signals are then obtained using one of the three reconstruction formulae replacing $V_{\tilde{f}}^{g}$ by $\overline{V}_{\tilde{f}}^{g}$. For the sake of simplicity, we will also denote these techniques by RT1, RT2 and RT3, and f_r a reconstructed signal using one of these methods.

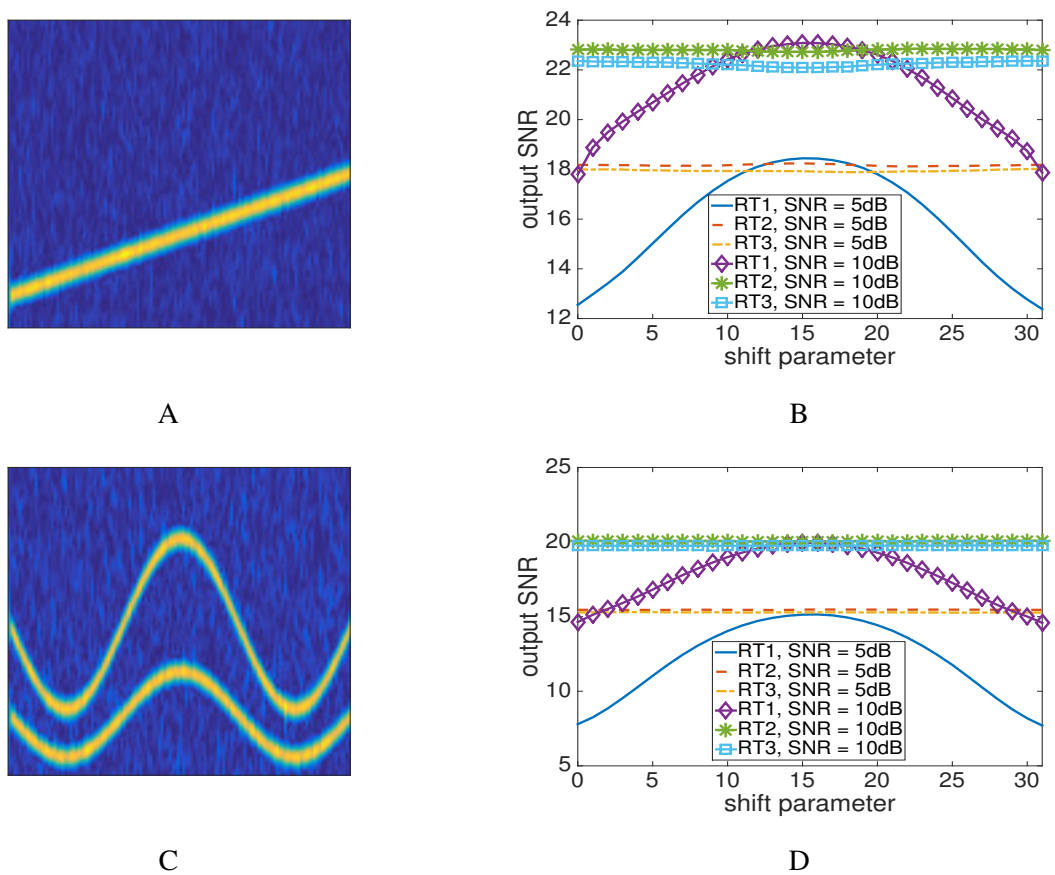


Fig. 1. A: STFT of a linear chirp; B: output SNR corresponding to the reconstruction of signal associated with STFT displayed in A, when the shift parameter varies, for different input SNR and for RT1, RT2, and RT3; C: STFT of a MCS; D: output SNR corresponding to the reconstruction of signal associated with STFT displayed in C, when the shift parameter varies, for different input SNRs and for RT1, RT2, and RT3.

To assess the quality of the reconstruction procedure using a shifted downsampled STFT, we remark that a good reconstruction technique should be very slightly sensitive to the shift parameter p. To measure

the quality of reconstruction we consider the output SNR corresponding to

$$SNR(f, f_r) = 20 \log_{10} \left(\frac{\|f\|_2}{\|f - f_r\|_2} \right). \tag{11}$$

Pointing out that the higher the SNR the lower the l_2 error $\|f - f_r\|_2$ we remark that, for RT1, we may write:

$$\sum_{q=0}^{R-1} |f[mR+q] - f_r[mR+q]|^2 = \sum_{q=0}^{R-1} \frac{1}{g[q-p]^2 N^2} | \sum_{k=0}^{N-1} \left(V_f^g(mR+p, \frac{k}{N}) - \overline{V}_{\hat{f}}^g(mR+p, \frac{k}{N} \right) e^{i2\pi \frac{k(q-p)}{N}} |^2. \tag{12}$$

If we define $K(p,q) := | \sum_{k=0}^{N-1} \left(V_f^g(mR+p, \frac{k}{N}) - \overline{V}_{\hat{f}}^g(mR+p, \frac{k}{N}) \right) e^{i2\pi \frac{k(q-p)}{N}} |^2$, and assume it is independent of p and q, then the l_2 error is minimal when $\sum_{q=0}^{R-1} \frac{1}{g[q-p]^2}$ is minimal for some p. For low-pass, positive, and symmetric windows (like Hamming or Bartlett windows, for instance), this corresponds to considering the central taps of filter g, i.e. $p = \lfloor R/2 \rfloor$. Regarding method RT2, energy conservation enables us to write:

$$\sum_{q \in \mathbb{Z}} |f[q] - f_r[q]|^2 = \sum_{m \in \mathbb{Z}} \frac{1}{N} \sum_{k=0}^{N-1} |V_f^g(mR+p, \frac{k}{N}) - \overline{V}_{\hat{f}}^g(mR+p, \frac{k}{N})|^2, \tag{13}$$

which is independent of p, provided $K(p,p)$ is independent of p.

To illustrate this, we consider a linear chirp defined for t in $\{0, ..., L-1\}$ by $f(t) = e^{2i\pi(0.05t + (0.10/L)t^2)}$, with $L = 4096$. The maximal reduced frequency is 0.25, and the Shannon-Nyquist sampling theory tells us that one needs at least $L/2$ samples to have perfect reconstruction, meaning the sampling period is half the critical one (i.e. the signal is oversampled by a factor of 2). Then, we compute the STFT of the signal using the Hamming window with length $L_g = 63$ and consider a number $N = 256$ of frequency bins, and $R = 32$ (the largest downsampling factor compatible with the three reconstruction techniques since $M = 32$). For $p = 0$, we obtain the representation of Figure 1 A, corresponding to a 256×128 matrix and, changing p does not change the matrix size. So, if one compares with the critical sampling, requiring 2048 values for f, the downsampling factor actually equals 16. Then, for different noise levels, shift parameter in $\{0, \cdots, R-1\}$ and input SNRs, we display in Figure 1 B, the output SNR corresponding to the reconstruction of the linear chirp using either RT1, RT2, or RT3.

For RT1, having numerically checked that $K(p,q)$ is almost independent of p and q, we notice that, as expected, best reconstruction results are obtained when $p = \lfloor R/2 \rfloor$. On the contrary, we check that RT2 is insensitive to p, as a consequence of energy conservation. The behavior of RT3 is very similar to that of RT2 but is more difficult to analyze. As we are interested in reconstructing the signal from its

downsampled STFT for any given shift parameter, RT1 should not be used. However, such a reconstruction should be borne in mind since it is used in the synchrosqueezing techniques discussed at the end of this paper.

It is also worth mentioning that when the SNR is equal to 10 dB, and no downsampling is applied, the output SNRs for the linear chirp and with either of the three methods RT1, RT2 and RT3 are 23.39, 23.27 and 22.87 dB respectively: looking at Figure 1 B, we notice that the denoising performance are not sensitive to downsampling factor since similar results are obtained using downsampled STFT in RT2, RT3 and also for RT1, provided the shift parameter is optimally chosen for the latter.

This behavior is also observable when applying the reconstruction techniques to a more general signal, as the MCS whose downsampled STFT is depicted in Figure 1 C, for which the output SNR is displayed in Figure 1 D. In this example, the reduced frequencies are $\frac{400}{L} + \frac{30}{L}\cos(3\pi\frac{t}{L})$ and $\frac{1000}{L} + \frac{60}{L}\cos(3\pi\frac{t}{L})$ with $L = 4096$, meaning that the maximal reduced frequency equals 0.38: the signal is over-sampled by a factor less than 2 (so with $R = 32$, the downsampling factor compared with the critical Nyquist sampling is even bigger than in the previous example). As previously, to use downsampled STFT instead of the full transform does not result in a significant performance loss. In the following section, we are going to exploit these findings to propose a novel algorithm for the reconstruction and denoising of the modes of a MCS.

IV. Mode Recontruction from Downsampled STFT of Noisy MCS

In the following, we study more in details MCSs f defined by:

$$f[n] = \sum_{l=1}^{K} f_l[n], \text{ with } f_l[n] = A_l[n]e^{i2\pi\phi_l[n]}, \tag{14}$$

where $A_l[n] > 0$ and $\phi_l'[n] > 0$. We also assume the f_ls, called modes or components, are separated with resolution Δ and with bounded modulation M, i.e. for all times n,

$$\begin{aligned} \forall\ 1 \leq l \leq K-1,\ \phi_{l+1}'[n] - \phi_l'[n] > 2\Delta \\ \forall\ 1 \leq l \leq K,\ |\phi_l''[n]| \leq M. \end{aligned} \tag{15}$$

Mode reconstruction in the TF plane is carried out by integrating STFT along the frequency axis and in the vicinity of so-called *ridges*, approximating the $\phi_l's$, called *instantaneous frequencies* (IF) of the modes. Therefore, we first explain how ridges are detected and then detail how mode retrieval is performed. In what follows, $\tilde{f}$ will be a noisy version of f as introduced in Section III.

A. Ridge Extraction from Noisy Downsampled STFT

We here recall how to compute an estimate of the ridges $(n, \phi_l'[n])_l$. Assuming knowledge of the number K of modes, we are going to use an adaption of the algorithm detailled in [2] or [1], which was originally proposed in [14]. This computes a local minimum of the functional :

$$E_{\tilde{f}}(\psi_1, \cdots, \psi_K) = \sum_{l=1}^{K} \sum_{n=0}^{L-1} -|V_{\tilde{f}}^{g}(n, \psi_l[n])|^2 + \lambda \psi_l'[n]^2 dt + \beta \psi_l''[n]^2, \tag{16}$$

so that ψ_l approximates ϕ_l'. However, equation (16) does not offer any algorithmic means to compute the ridges and does not take into account the downsampling factor. It was shown in [8] through a numerical study, that to use regularization terms in the energy functional resulted in less accurate ridge estimation. Thus, to detect the ridge of a monocomponent signal and in the downsampled STFT context, (16) reduces to:

$$\psi[mR + p] = \underset{0 \leq k \leq N-1}{\text{argmax}} \, |V_{\tilde{f}}^{g}(mR + p, \frac{k}{N})|, \tag{17}$$

making m vary. The performance of this IF estimator is available in [15].

Algorithm 1 (Ridge extraction from downsampled STFT)

Pick $p \in \{0, \cdots, R-1\}$

Define $Z_{p,R} = \{mR + p, m \in \lfloor \frac{L-1}{R} \rfloor\}$

$\Delta_f = R \times M \times N$

for $l = 1$ to K **do**

1. Pick $t_m = mR + p \in Z_{p,R}$.
2. $p_{m,l} := \underset{k}{\text{argmax}} \, |V_{\tilde{f}}^{g}(t_m, \frac{k}{N})|$.
3. $I_{m,l} := [p_{m,l} - \Delta_f, p_{m,l} + \Delta_f] \bigcap \{0, ..., N-1\}$.
4. $p_{m-1,l} := \underset{k \in I_{m,l}}{\text{argmax}} \, |V_{\tilde{f}}^{g}(t_{m-1}, \frac{k}{N})|$, iterate backward.
5. $p_{m+1,l} := \underset{k \in I_{m,l}}{\text{argmax}} \, |V_{\tilde{f}}^{g}(t_{m+1}, \frac{k}{N})|$, iterate forward.
6. For $t_m \in Z_{p,R}$, set $V_{\tilde{f}}^{g}(t_m, k/N) := 0$,
   for $k \in [p_{m,l} - \Delta N, p_{m,l} + \Delta N]$

Sort $(p_{m,l})_l$ in increasing order to obtain $(\tilde{p}_{m,l})_l$

$(\psi_l[t_m])_m = (\frac{\tilde{p}_{m,l}}{N})_m$

To extract the ridges of a MCS from its downsampled STFT, we generalize (17), using a procedure summarized in Algorithm 1. The ridges are extracted one by one as follows. One first picks a random time

and finds the local maximum of the modulus of STFT at that time (steps 1 and 2, this point is supposed to be on a ridge), then, since mode modulation is bounded by M, one finds neighboring points on the ridge using the fact that the maximum displacement of a point on a ridge between two successive locations is $\pm M \times R$ (which means $\pm M \times R \times N$ in terms of indices, steps 3–5), and finally iterates backward and forward (steps 4 and 5). One then removes the detected ridge and corresponding TF neighborhood (step 6, the ridges are associated with K sufficiently separated local maxima at each time instant), and proceeds with the remaining ridges.

Note that to improve the quality of ridge detection, for each mode, steps 1 to 5 are iterated and the retained ridge corresponds to the most energetic one after these iterations. This algorithm is similar to that proposed in [8] for ridge extraction, except that we take into account the downsampling factor R in the search for $p_{m-1,l}$ and $p_{m+1,l}$ from $p_{m,l}$.

B. Mode Reconstruction

Based on RT2 and the signal denoising procedure from downsampled STFT introduced in Section III, we define a novel technique for the retrieval of the modes of a MCS (to consider RT3 for mode reconstruction instead would not change the results that much). Let us focus on the reconstruction of the l-th mode, assuming ψ_l is known. For that purpose, we recall that a first order approximation of STFT close to the lth ridge and in the continuous time setting, reads [3]:

$$V^g_{\tilde{f},c}(t,\omega) \approx A_l(t)\hat{g}(\omega - \phi_l'(t)), \tag{18}$$

where $V^g_{f,c}(t,\omega) = \int_{\mathbb{R}} f(u)g(u-t)e^{-i2\pi\omega(u-t)}du$ (assuming g is a real window) and $\hat{f}(\omega) = \int_{\mathbb{R}} f(u)e^{-i2\pi\omega u}du$. If the window g is the Gaussian window $\sigma_s^{-1/2}e^{-\pi\frac{x^2}{\sigma_s^2}}$ then $\hat{g}(\xi) = \sigma_s^{1/2}e^{-\pi\sigma_s^2\xi^2}$. Using such a window, considering the approximation of ϕ_l' by ψ_l, and adapting (18) to our discrete time and frequency settings, we get at time $t_m := mR + p$:

$$V^g_{\tilde{f}}(t_m, \frac{k}{N}) \approx |V^g_{\tilde{f}}(t_m, \psi_l[t_m])|e^{-\pi\sigma_s^2(\frac{k}{N}-\psi_l[t_m])^2},$$

for $\frac{k}{N}$ sufficiently close to $\psi_l[t_m]$. If, as previously, the noise variance is σ^2, the coefficients associated with mode l one selects at each time t_m, following the procedure detailled in Section III, correspond to the indices k satisfying:

$$|V^g_{\tilde{f}}(t_m, \psi_l[t_m])|e^{-\pi\sigma_s^2(\frac{k}{N}-\psi_l[t_m])^2} \geq 3\sqrt{2}\sigma\|g\|_2,$$

namely the indices inside the interval $J_{l,m} := [N\psi_l[t_m] - \Gamma_{l,m}, N\psi_l[t_m] + \Gamma_{l,m}]$, with

$$\Gamma_{l,m} = \frac{N}{\sigma_s}\sqrt{-\frac{1}{\pi}\log\left(\frac{3\sqrt{2}\sigma\|g\|_2}{|V_{\tilde{f}}^g(t_m,\psi_k[t_m])|}\right)}.$$

Knowing interval $J_{l,m}$, mode f_l is reconstructed through:

$$f_{r,l}[n] := \sum_{m\in\mathbb{Z}}\sum_{k\in J_{l,m}} V_{\tilde{f}}^g(t_m,\frac{k}{N})g[n-t_m]\frac{e^{i2\pi\frac{k(n-t_m)}{N}}}{N}. \tag{19}$$

This method will be denoted by M_1 in the sequel. Doing so, one selects the portion of signal above noise level in the vicinity of each ridge. However, such a simple expression for $J_{l,m}$ exists only because one assumes the window is Gaussian. Also, the width of interval $J_{l,m}$ only depends on the magnitude of STFT at local maxima location, but not on the frequency modulation of the corresponding modes. Therefore, an alternative to compute the frequency interval for mode reconstruction is to define $\eta_{l,m}^{[1]}$ and $\eta_{l,m}^{[2]}$:

$$\eta_{l,m}^{[1]} = \underset{k}{\operatorname{argmax}}\left\{\frac{k}{N} < \psi_l[t_m], |V_{\tilde{f}}^g(t_m,\frac{k}{N})| < 3\sqrt{2}\sigma\|g\|_2\right\}$$
$$\eta_{l,m}^{[2]} = \underset{k}{\operatorname{argmin}}\left\{\frac{k}{N} > \psi_l[t_m], |V_{\tilde{f}}^g(t_m,\frac{k}{N})| < 3\sqrt{2}\sigma\|g\|_2\right\},$$

and then set :

$$J_{l,m} := \left[N\psi_l[t_m] - \eta_{l,m}^{[1]}, N\psi_l[t_m] + \eta_{l,m}^{[2]}\right]. \tag{20}$$

In such a case, the interval is no longer symmetric with respect to the maximum amplitude location and adapts to frequency modulation. This method will be denoted by M_2 in the sequel. The comparison between methods M_1 and M_2 will be carried out in the Section Numerical Applications.

V. Parameters estimation

A. Estimation of the Threshold for Denoising and Mode Reconstruction

To perform coefficients selection for STFT denoising prior to mode reconstruction, one needs to estimate $\gamma := \sigma\|g\|_2$ from the STFT of $\tilde{f}$. This is done by considering the *median absolute deviation* approach [16], which we adapt to our discrete STFT context. Indeed, as shown in [17] but for wavelet decompositions, since there are very few STFT coefficients associated with the signal and since the real part of STFT of the noise is a zero mean Gaussian variable, γ can be estimated through:

$$\hat{\gamma}_p = \frac{\text{median}\,|\Re\left(V_{\tilde{f}}^g(mR+p,\frac{k}{N})\right)_{m,k}|}{0.6745}. \tag{21}$$

Computing the estimation $\hat{\gamma}_p$ given by formula (21) on the linear chirp associated with Figure 1 A, when the SNR equals 10 dB, we get an estimate fluctuating around 1.22 (the ground truth being 1.14), the amplitude of the variations are less than 1% when both the shift parameter p and the downsampling factor R vary (tested values for R were 8, 16, 32 and 64, while $0 \leq p \leq R-1$). Similar conclusions could be obtained with another SNR. Finally, the same type of results are derived with more complicated signals like the one associated with Figure 1 C.

B. Optimal Window Length Determination

An important parameter in all these TF methods is the length of the analysis window. Our approach is to use, as in [8] [18], the Rényi entropy to determine an optimal window length. For the Gaussian window, this corresponds to the determination of an optimal window parameter σ_s, and an optimal L_g for the Hamming window. In our framework, the optimal window length is the one leading to the most concentrated representation, i.e. associated with the lowest Rényi entropy defined by:

$$H(\sigma) = \frac{1}{1-\alpha} \log_2 \left(\sum_{k=0}^{N-1} \sum_{n=0}^{L-1} \left(\frac{|V_{\tilde{f}}^{g}(n, \frac{k}{N})|}{\|V_{\tilde{f}}^{g}\|_1} \right)^{\alpha} \right). \tag{22}$$

Here, we measure the Rényi entropy on the full STFT, but the conclusions would be the same on a downsampled STFT. Parameter α is set to 3 but another value would lead to similar results. In Figure 2

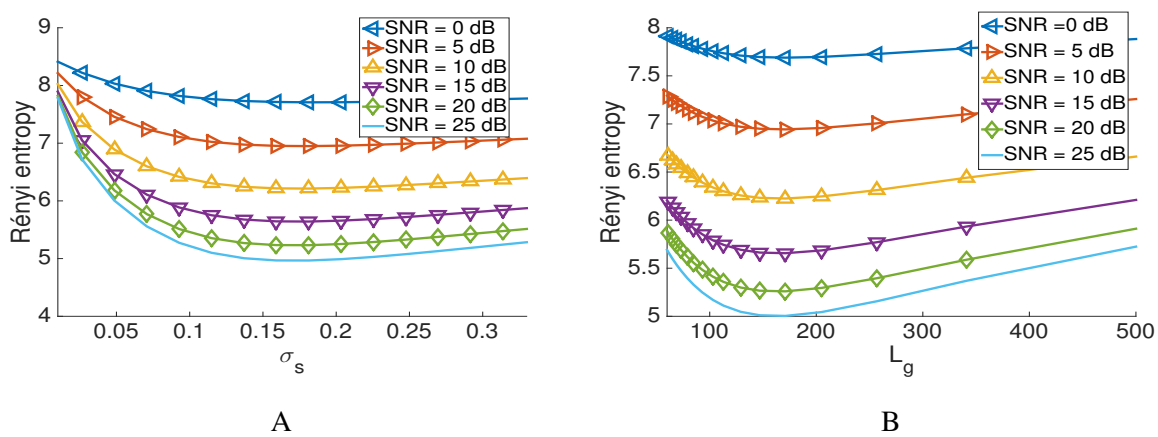


Fig. 2. A: Computation of Rényi entropy with respect to σ_s when the Gaussian window is used to compute STFT, for different noise level ; B: same as A but when the Hamming window is used instead

A and B, we display the Rényi entropy of the STFT of the signal associated with Figure 1 C, for various input SNRs, and with respect to the filter length parameter, for the Gaussian and Hamming windows

respectively. Looking at these figures, and as already noticed in [8], the minimum entropy clearly puts forward an optimal window length which varies very little with the noise level. In these simulations, we took $N = 512$ and then the optimal window parameter σ_s equals 0.15 and the optimal L_g equals 161 for the Hamming window.

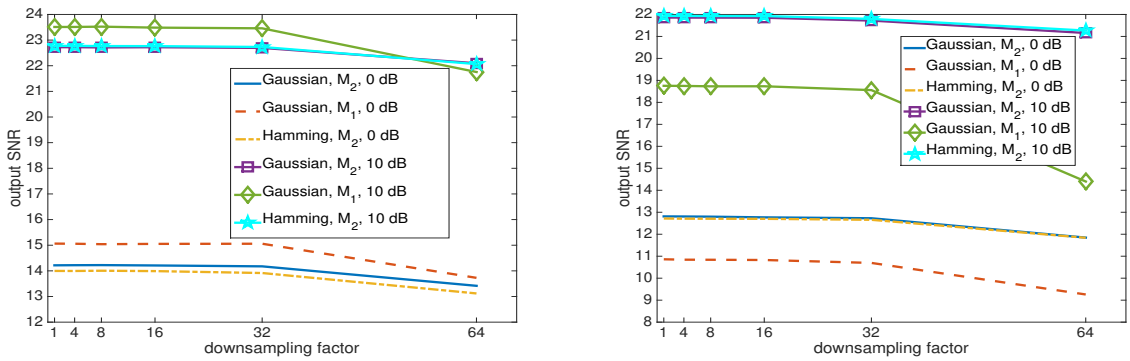
VI. Numerical Applications

In this section, we first investigate the quality of mode reconstruction with respect to the type of reconstruction method, i.e. M_1 or M_2, type of filter, downsampling factor and shift parameter. Then, we compare the proposed mode reconstruction methods with other ones based on synchrosqueezing transform [6] [8]. In this regard, a particular emphasis is put on the amount of information needed to perform mode reconstruction (the MATLAB code to reproduce the different figures of this paper is available at https://github.com/meignen/STFT_MCS_DENOISING).

A. Sensitivity of Mode Reconstruction to Filter Choice, Downsampling Factor and Shift Parameter

We here study the sensitivity of mode reconstruction techniques M_1 and M_2, to the choice of downsampling factor, shift parameter, and filter g. Our experiments are again based on noisy versions of the signal associated with Figure 1 C, with $N = 512$, and the optimal filter is chosen as explained in the previous subsection (we only consider Hamming or Gaussian filter in our simulations). Taken into account the optimal length associated with each of these filters, the largest downsampling factor one can consider is $R = 128$, but to take $R \leq 64$ is sufficiently informative on the behaviors of the proposed techniques. To measure the performance of the different reconstruction techniques we compute the output SNRs associated with the reconstruction of each mode, i.e. $SNR(f_l, f_{r,l})$, where the output SNR is defined in (11) and $f_{r,l}$ in (19).

As for the whole signal reconstruction procedure studied in section III, the output SNR, when considering mode reconstruction, is almost insensitive to the shift parameter. Therefore, for the sake of simplicity, we omit this parameter in the simulations and consider $p = 0$. The results depicted in Figure 3 A, for an input SNR equal to 0 dB or 10dB, show that the quality of reconstruction for f_1 worsens when the downsampling factor increases, whatever the type of reconstruction method used, but the performance loss is not significant at least for M_2. On that example, M_1 behaves slightly better than M_2 but the discrepancy between the results provided by the two methods remains within a range of 1 dB. Finally, to use the Hamming or the Gaussian window makes very little difference.



A B

Fig. 3. A: output SNR associated with the reconstruction of mode f_1 of the signal associated with Figure 1 C, for different Rs, different optimized filters, for either reconstruction technique M_1 or M_2, and when the input SNR is 0 dB or 10 dB; B: same as A but for mode f_2

If one carries out the same study on mode f_2, which is much more modulated than f_1, the importance of taking into account the modulation in the reconstruction procedure transpires. The gain obtained by using M_2 rather than M_1 is at least 2 dB at high noise level, and the benefits of using the former technique is even greater when the noise level is low. Based on these results, we will use M_2 in the comparison with synchrosqueezing techniques that follows.

B. Comparison with Synchrosqueezing Transform Based Methods

Recently, a technique extending *STFT-based synchrosqueezing transform* (FSST) to MCSs containing strongly frequency modulated modes was proposed in [6] [19], and was called *second order synchrosqueezing* (VSST) (original description of STFT-based synchrosqueezing transform can be found in [3] [20]). Such a method was also used in a demodulation algorithm, still for the purpose of mode retrieval [8]. After having recalled the principle of VSST and of the algorithm based on VSST and demodulation for mode retrieval [8], we are going to carry out comparisons between these techniques and M_2.

The principle of VSST is to reassign the TF representation given by STFT, along the frequency axis, using the following complex estimate of the second order derivative of the phase of $\tilde{f}$, defined as follows:

$$\tilde{q}_{\tilde{f}}(n,\frac{k}{N}) = \frac{1}{2i\pi}\frac{V_{\tilde{f}}^{g''}(n,\frac{k}{N})V_{\tilde{f}}^{g}(n,\frac{k}{N}) - (V_{\tilde{f}}^{g'}(n,\frac{k}{N}))^2}{V_{\tilde{f}}^{tg}(n,\frac{k}{N})V_{\tilde{f}}^{g'}(n,\frac{k}{N}) - V_{\tilde{f}}^{tg'}(n,\frac{k}{N})V_{\tilde{f}}^{g}(n,\frac{k}{N})}.$$

Introducing $\tilde{\omega}_{\tilde{f}}(n, \frac{k}{N}) = \frac{\partial_t V_{\tilde{f}}^g(n, \frac{k}{N})}{2i\pi V_{\tilde{f}}^g(n, \frac{k}{N})}$ and $\tilde{t}_{\tilde{f}}(n, \frac{k}{N}) = n - \frac{\partial_\eta V_{\tilde{f}}^g(n, \frac{k}{N})}{2i\pi V_{\tilde{f}}^g(n, \frac{k}{N})}$, enables the definition of an IF estimate of $\tilde{f}$ as [19]:

$$\hat{\omega}_{\tilde{f}}^{(2)}(n, \frac{k}{N}) = \begin{cases} \Re\{\tilde{\omega}_{\tilde{f}}(n, \frac{k}{N}) + \tilde{q}_{\tilde{f}}(n, \frac{k}{N})(n - \tilde{t}_{\tilde{f}}(n, \frac{k}{N}))\} \\ \qquad\qquad \text{if } \partial_t \tilde{t}_{\tilde{f}}(n, \frac{k}{N}) \neq 0 \\ \Re\{\tilde{\omega}_{\tilde{f}}(n, \frac{k}{N})\} \text{ otherwise,} \end{cases}$$

which is proved to be exact for a noise-free Gaussian modulated linear chirp [19]. VSST then consists of reassigning the STFT of f through:

$$TV_{\tilde{f}}^g(n, \frac{k}{N}) = \frac{1}{g[0]} \sum_{\substack{|V_{\tilde{f}}^g(n, \frac{q}{N})| \geq \gamma \\ \text{round}(N\hat{\omega}_{\tilde{f}}^{(2)}(n, \frac{q}{N})) = k}} V_{\tilde{f}}^g(n, \frac{q}{N}), \tag{23}$$

where round(X) denotes the nearest integer to X and γ is the threshold introduced in Section III. Note that (23) can be viewed as a local version of RT1 (with $R = 1$) , the coefficient being summed up at their reallocation point.

The reconstruction of f_l is subsequently performed by means of the following formula:

$$f_l[n] \approx \sum_{|k - \text{round}(N\phi_l'[n])| < d} TV_{\tilde{f}}^g(n, \frac{k}{N}). \tag{24}$$

We shall remark that d is an integer that enables to compensate for reassignment and ridge estimation errors. It can be increased up to Δ and, in the absence of noise, and, if one uses a large d, one ends up with a reconstruction close to RT1 (except that one performs the reconstruction mode by mode) so that output SNR is better. In noisy situations, this appears to be numerically also the case since most of the noise has been removed by the thresholding procedure. In practice ϕ_l' is unknown and is replaced by its approximation ψ_l computed on $TV_{\tilde{f}}^g$, using the same approach as the one developed in Algorithm 1 but with $R = 1$.

Alternatively, one can use a demodulation algorithm prior to reconstruction as proposed in [8], by using the ridge estimate given by $(\psi_l)_{l=1,\cdots,K}$, and reconstruct the modes through Algorithm 2 (in which b_0 is a constant and Ψ_l a primitive of ψ_l null at 0). This method was shown, in [8], to behave better in terms of mode retrieval.

At first glance, since with VSST techniques, the information is reassigned in the vicinity of the ridge associated with one mode, for a given time, mode reconstruction involves much less coefficients than with STFT [6]. Nevertheless, we have just proven that we could still perform mode reconstruction from a downsampled version of STFT which is not the case with VSST. So, even though VSST leads to a much

Algorithm 2 (mode reconstruction using demodulation and VSST)

Compute ridge estimates $(\psi_1, \cdots, \psi_K)$ from VSST of $\tilde{f}$

for $l = 1$ to K **do**

1. Compute $\tilde{f}_{D,l}[n] = \tilde{f}[n]e^{-i2\pi(\Psi_l[n]-b_0 n)}$
2. Extract ridge $\psi_{D,l}$ from VSST of $\tilde{f}_{D,l}$ using single ridge detection in $[b_0 - \Delta, b_0 + \Delta]$.
3. Reconstruct f_l through:

$$f_l[n] \approx \left(\sum_{|k-N\psi_{D,l}(t)|<d} TV^g_{\tilde{f}_{D,l}}(n, \frac{k}{N}) \right) e^{i2\pi(\Psi_l[n]-b_0 n)}.$$

sharper representation than STFT, if one is interested in signal reconstruction using the least amount of information as possible, we are going to show that the reconstruction based on STFT performs better than those based on VSST.

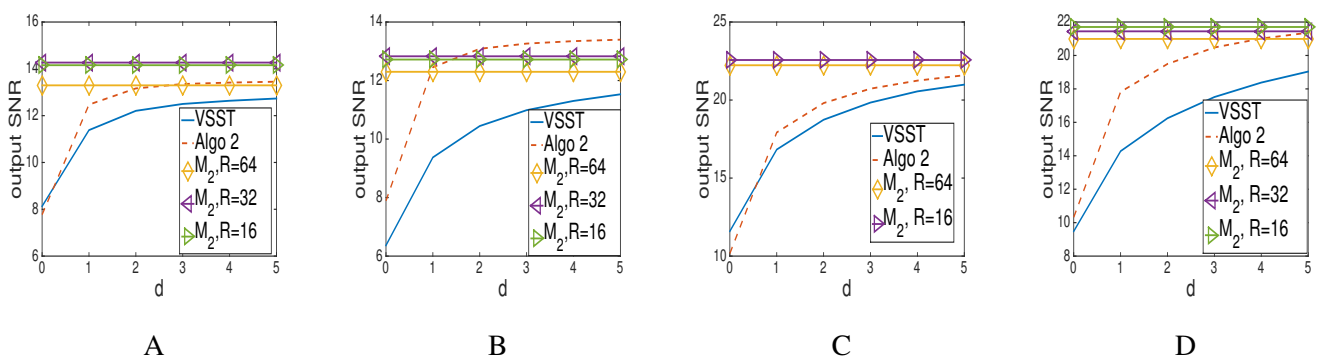


Fig. 4. A: output SNR associated with the reconstruction of mode f_1 of the signal associated with Figure 1 C, using different VSST techniques and with respect to parameter d, or M_2 with optimized Hamming filter and different downsampling factors R (SNR 0 dB); B: same as A but for mode f_2; C: same as A but with SNR 10 dB; D: same as B but with SNR 10 dB.

To quantify this, we compute the mode reconstruction performance for the signal associated with Figure 1 C. We investigate the behavior of the just recalled VSST techniques (either based on direct reconstruction or using demodulation beforehand) with respect to parameter d, along with mode reconstruction based on method M_2, using the optimized Hamming filter and when the downsampling parameter R varies. The output SNRs are displayed in Figure 4 A-D, and correspond to input SNRs 0 or 10 dB (the results are averaged over 5 realizations, Algo 2 denoting the reconstruction technique associated with Algorithm 2). As the reconstruction with M_2 does not depend on d, in these figures, the mode reconstruction results

with this technique are displayed as constant functions. As already noticed, M_2 is not much sensitive to the downsampling factor when the latter goes from $R = 16$ to $R = 64$, for both f_1 and f_2.

Analyzing the results regarding VSST techniques, and as reported in [8], to demodulate the signal as proposed in Algorithm 2, results in better mode reconstruction than when using VSST directly. Now comparing M_2 with VSST based techniques, M_2 leads better results when only the information on the ridge is considered for VSST ($d = 0$), but when d increases, typically $d = 5$, mode reconstruction performed with Algorithm 2 seems to perform similarly to M_2 in most cases.

However, there is one key issue which is hidden in the just computed output SNRs, which is the number of coefficients used to perform the reconstruction with each method. Indeed, for a fair comparison, we compute the number of coefficients required by the different reconstruction procedures. If the reconstruction procedure is based on Algorithm 2 or on VSST, the total number of coefficients to be stored is $N_{VSST} = K(2d+1)L$. On the contrary, the number of coefficients used by M_2 corresponds to:

$$N_{M_2} = \sum_{l=1}^{K} \sum_{m} \# \left\{ J_{l,m} \bigcap \mathbb{Z} \right\}, \tag{25}$$

where $J_{l,m}$ is the interval defined in (20) and $\#\{X\}$ the cardinal of set X. Note that the range for m depends on the downsampling factor R. In Figure 5, we depict for input SNRs equal to 0 or 10 dB, the output SNRs associated with mode reconstruction with respect to the number of coefficients kept, i.e. N_{VSST} or N_{M_2}, divided by signal length L. Doing so, we highlight the very different nature of the mode reconstruction techniques based on Algorithm 2 and M_2. Indeed, whatever the type of modes considered, for a given output SNR, mode retrieval is performed using much less information using the latter technique than the former. But more than that: the latter technique, while denoising the modes, can also compress the whole signal, since the number of coefficients kept over the signal length L is lesser than one when $R \geq 64$. On the contrary, VSST based denoising techniques always require more than L coefficients. In this example, the amount of information needed to get the same output SNR is about 15 times larger with the technique associated with Algorithm 2 than with M_2. Note finally that for the former technique the results are not sensitive to the modulation (f_1 reconstruction versus that of f_2), which is expected because this algorithm demodulates each mode before proceeding with reconstruction (see step 1 of Algorithm 2).

C. Application to Real Data

In this section, we investigate the behavior of our new method for the denoising of a bat echolocation call, whose noise-free STFT is depicted in Figure 6 A. As one does not know the ground truth in terms

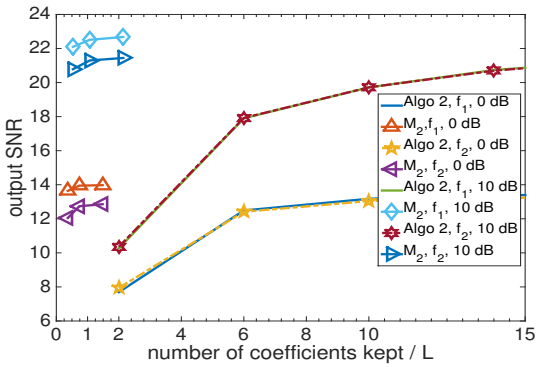


Fig. 5. Output SNR associated with the reconstruction of the signal associated with Figure 1 C with respect to the number of coefficients kept divided by the signal length (input SNR 0 or 10 dB)

of the modes to be extracted, one cannot compute output SNR for each mode ; the only thing one can do is to compute a global output SNR, as defined in (11). In this regard, we are going to check that we do not lose much information by considering mode extraction as in M_2. Another difference between this signal and the one associated with Figure 1 C is that the modes only last for a certain amount of time, which makes ridge extraction more complicated. For denoising purpose, we thus investigate three different strategies: we either use M_2, the technique based on Algorithm 2, or the denoising procedure introduced in Section III, which does not perform ridge extraction before denoising.

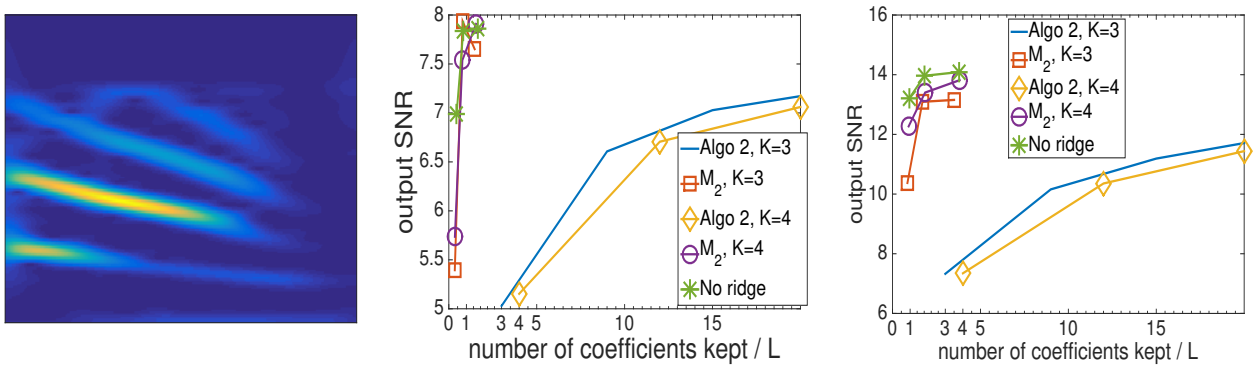


Fig. 6. A: STFT of a bat echolocation call, B: output SNR associated with the reconstruction of the signal associated with A with respect to the number of coefficients kept divided by the signal length (input SNR 0 dB), the sensitivity to the number K of modes is also tested; C: same as B but with an input SNR of 10 dB.

We again analyze the quality of denoising with respect to the number of coefficient kept over the length of the signal, the results depicted in Figure 6 B and C, respectively correspond to a noise level of 0 and 10 dB. The number K of modes is set to 3 or 4 in M_2 and also in Algorithm 2 . We first remark that while the latter technique is sensitive to the number of modes (N_{Vsst} is linear with respect to K), M_2 is not. Indeed, if one considers an irrelevant mode number with M_2, the corresponding coefficients will be mostly below the noise level and will thus not be considered. Another interesting fact is that, in such a context, to extract the mode for denoising purpose may be useless, because the coefficients above the noise threshold are almost all located in the vicinity of the different ridges: given a downsampling factor, the number of selected coefficients is almost the same with M_2 or with the denoising procedure introduced in Section III (denoted by "No ridge" in Figures 6 B and C). Finally, we also remark that when the downsampling factor is large, the denoising performance becomes worse when extracting the ridges in the denoising process: if one wants to denoise the signal while compressing it, in such a situation, ridge extraction should be precluded.

VII. Conclusion

In this paper, we first recalled different signal reconstruction procedure using downsampled versions of short-time Fourier transform. From this, we have derived a new algorithm for signal denoising which we have then adapted to the denoising of multicomponent signals and mode extraction. We then showed that the proposed new technique compares favorably, on simulated and real data, with the most recent denoising techniques based on synchrosqueezing in terms of the number of coefficients used in the reconstruction procedure. Future work should involve the generalization of the proposed denoising approach to non Gaussian noises.

Appendix

If g is such that, for any n, $\sum_{m\in\mathbb{Z}} g[n-mR]^2 = 1$, from (2) and putting $q = n + mR + p - M$, we may write:

$$f[q]g[q-mR-p] = \frac{1}{N}\sum_{k=0}^{N-1} V_f^g(mR+p, \frac{k}{N})e^{i2\pi\frac{k(q-mR-p)}{N}}.$$

From this we deduce:

$$f[q]\sum_{m\in\mathbb{Z}} g[q-mR-p]^2 = \sum_{m\in\mathbb{Z}}\frac{1}{N}\sum_{k=0}^{N-1} V_f^g(mR+p,\frac{k}{N})e^{i2\pi\frac{k(q-mR-p)}{N}}g[q-mR-p]$$

$$\Leftrightarrow f[q] = \frac{\sum\limits_{m\in\mathbb{Z}}\frac{1}{N}\sum\limits_{k=0}^{N-1} V_f^g(mR+p,\frac{k}{N})e^{i2\pi\frac{k(q-mR-p)}{N}}g[q-mR-p]}{\sum\limits_{m\in\mathbb{Z}} g[q-mR-p]^2}.$$

If $\sum\limits_{m\in\mathbb{Z}} g[n-mR]^2 = 1$, we get (4), and if it is only non-zero for any n, (6) follows.

Regarding energy conservation, from (2) we may write:

$$\sum_{m\in\mathbb{Z}}\sum_{q=mR+p-M}^{N-1+mR+p-M} f[q]^2 g[q-mR-p]^2 = \sum_{m\in\mathbb{Z}}\sum_{q=mR+p-M}^{N-1+mR+p-M} |\frac{1}{N}\sum_{k=0}^{N-1} V_f^g(mR+p,\frac{k}{N})e^{i2\pi\frac{k(q-mR-p)}{N}}|^2$$

$$\Leftrightarrow \sum_{m\in\mathbb{Z}}\sum_{q\in\mathbb{Z}} f[q]^2 g[q-mR-p]^2 = \sum_{m\in\mathbb{Z}}\frac{1}{N}\sum_{k=0}^{N-1}|V_f^g(mR+p,\frac{k}{N})|^2$$

$$\Leftrightarrow \sum_{q\in\mathbb{Z}} f[q]^2 = \sum_{m\in\mathbb{Z}}\frac{1}{N}\sum_{k=0}^{N-1}|V_f^g(mR+p,\frac{k}{N})|^2.$$

The proof of (7) is similar to that of (4). Indeed, we have:

$$f[q]\sum_{m\in\mathbb{Z}} g[q-mR-p] = \sum_{m\in\mathbb{Z}}\frac{1}{N}\sum_{k=0}^{N-1} V_f^g(mR+p,\frac{k}{N})e^{i2\pi\frac{k(q-mR-p)}{N}}$$

$$\Leftrightarrow f[q] = \frac{\sum\limits_{m\in\mathbb{Z}}\frac{1}{N}\sum\limits_{k=0}^{N-1} V_f^g(mR+p,\frac{k}{N})e^{i2\pi\frac{k(q-mR-p)}{N}}}{\sum\limits_{m\in\mathbb{Z}} g[q-mR-p]}.$$

If $\sum\limits_{m\in\mathbb{Z}} g[n-mR] = 1$ for any n, then one obtains (7) and otherwise if the sum is only non-zero (8) follows.

REFERENCES

[1] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Sig. Proc.*, vol. 43, no. 5, pp. 1068–1089, 1995.

[2] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 243–261, 2011.

[3] T. Oberlin, S. Meignen, and V. Perrier, "The Fourier-based synchrosqueezing transform," in *IEEE Int. Conf. Acoust., Speech., Sig. Proc. (ICASSP)*, 2014, pp. 315–319.

[4] H.-T. Wu, "Adaptive analysis of complex data sets," *PhD, Princeton University*, 2012.

[5] S. Wang, X. Chen, G. Cai, B. Chen, X. Li, and Z. He, "Matching demodulation transform and synchrosqueezing transform in time-frequency analysis," *IEEE Trans. Sig. Proc.*, vol. 62, no. 1, pp. 69–84, 2014.

[6] T. Oberlin, S. Meignen, and V. Perrier, "Second-order synchrosqueezing transform or invertible reassignment? Towards ideal time-frequency representations," *IEEE Trans. Sig. Proc.*, vol. 63, no. 5, pp. 1335–1344, March 2015.

[7] D.-H. Pham and S. Meignen, "High-order synchrosqueezing transform for multicomponent signals analysiswith an application to gravitational-wave signal," *IEEE Trans. Sig. Proc.*, vol. 65, no. 12, pp. 3168–3178, 2017.

[8] S. Meignen, D.-H. Pham, and S. McLaughlin, "On demodulation, ridge detection and synchrosqueezing for multicomponent signals," *IEEE Trans. Sig. Proc.*, vol. 65, no. 8, pp. 2093–2103, 2017.

[9] H. Yang, "Statistical analysis of synchrosqueezed transforms," *Appl. Comput. Harm. Anal.*, vol. doi.org/10.1016/j.acha.2017.01.001, 2017.

[10] J. Xiao and P. Flandrin, "Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement," *IEEE Trans. Sig. Proc.*, vol. 55, no. 6, pp. 2851–2860, 2007.

[11] I. Daubechies, Y. Wang, and H.-T. Wu, "ConceFT: concentration of frequency and time via multitapered synchrosqueezed transform," *Phil. Trans. of the Roy. Soc. Lond. A: Math., Phys. and Eng. Sci.*, vol. 2065, no. 374, 2016.

[12] D. W. Griffin and S. L. Jae, "Signal estimation from modified short-time Fcourier transform," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol. 32, no. 2, pp. 236–243, 1984.

[13] J. B. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol. 25, pp. 235–238, 1977.

[14] R. Carmona, W. Hwang, and B. Torresani, "Characterization of signals by the ridges of their wavelet transforms," *IEEE Trans. Sig. Proc.*, vol. 45, no. 10, pp. 2586–2590, 1997.

[15] L. Stankovic, M. Dakovic, and V. Ivanovic, "Performance of spectrogram as if estimator," *Electron. Lett.*, vol. 37, no. 12, pp. 797–799, 2001.

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

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

[18] R. Baraniuk, P. Flandrin, A. Janssen, and O. Michel, "Measuring time-frequency information content using the renyi entropies," *IEEE Trans. Info. Theo.*, vol. 47, no. 4, pp. 1391–1409, 2001.

[19] R. Behera, S. Meignen, and T. Oberlin, "Theoretical analysis of the second-order synchrosqueezing transform," *Appl. Comput. Harm. Anal.*, vol. , doi:10.1016/j.acha.2016.11.001, 2016.

[20] G. Thakur and H.-T. Wu, "Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples." *SIAM J. Math. Anal.*, vol. 43, no. 5, pp. 2078–2095, 2011.