On Local Chirp Rate Estimation in Noisy Multicomponent Signals: With an Application to Mode Reconstruction

Nils Laurent, Marcello A Colominas, Sylvain Meignen

On Local Chirp Rate Estimation in Noisy Multicomponent Signals: With an Application to Mode Reconstruction

N. Laurent, M. A. Colominas, S. Meignen

***Abstract*—In this paper, our goal is to investigate local chirp rate (CR) estimation in noisy multicomponent signals. The focus is put on improving a specific type of local CR estimator used in the second order synchrosqueezing transform, which proves to be very inaccurate in the presence of noise. More precisely, the noise creates spurious oscillations in the local CR estimate, and we first put the emphasis on the terms responsible for these oscillations. Then we propose a novel technique to filter them out, resulting in a much more accurate local CR estimator. We finally show how to use the latter to improve mode reconstruction and investigate in what way the new CR estimator is useful in the context of voice signals.**

***Index Terms*—Time-frequency analysis, Fourier-based synchrosqueezing transform, reassignment methods**

I. Introduction

TIME-FREQUENCY (TF) analysis is a very dynamic field of signal processing, since it enables to process non-stationary signals [1], [2] encountered in many fields such as audio (speech and music), or biomedicine (electrocardiogram, thoracic, and abdominal movement signals) [3], [4]. These signals are often modeled as *multicomponent signals* (MCSs), namely the sum of AM/FM components (or modes) and *time-frequency representations* (TFRs) such as the *continuous wavelet transform* [5]–[7] or the *short-time Fourier transform* (STFT) [8] are commonly used for their study. The main reason why such TFRs are so popular for studying MCSs is that each mode can be associated with a TF curve, called *ridge*, corresponding to local modulus maxima of the TFR along the frequency axis. In particular, many mode reconstruction techniques are based on the analysis of these TFRs in the vicinity of these ridges [9], [10].

However, these TFRs are hampered by the choice for the window or wavelet which is constrained by the Heisenberg-Gabor uncertainty principle [11]: a very short temporal window in STFT or a wavelet associated with a good time localization in CWT both lead to a bad frequency localization, and vice-versa. Many works in the past decades have tackled this limitation, by using, for instance, quadratic TFRs, e.g. Wigner-Ville distributions [2], which are not constrained by the uncertainty principle, but exhibit strong interference hampering the representation. Another technique, called the *reassignment method* (RM) [12], proposed another means to improve the readability of the TFR by reassigning its coefficients, the reassigned representation being however not invertible.

To improve the readability of the continuous wavelet transform of an MCS, an invertible reassignment process called *synchrosqueezing transform* was introduced [13], [14], and then adapted to STFT to obtain the so-called *STFT-based SST* [15]–[18]. As synchrosqueezing transforms in their initial formulation are only efficient on signals made of quasi-harmonic components, variants were proposed to deal with modes containing local linear chirp-like frequency modulations, through the so-called *second order synchrosqueezing transform* in the STFT framework [18], [19] and also in the wavelet framework [20]. Synchrosqueezing transform was finally generalized to modes with fast oscillating phases, through the *higher order synchrosqueezing transform* [21].

One of the specificity of second and higher order synchrosqueezing transforms is that they are based on a local *chirp rate* (CR) estimator, which is computed in a non-parametric way, CR being yet another term for the derivative of the frequency. CR estimation is very important in the domain of *synthetic aperture radar* and *inverse synthetic aperture radar imaging* [22], since, due to the motion of a target, the radar return signals are usually chirps, and their CRs include the information about the target, such as the location and the velocity. Therefore, CR estimation is critically important in these applications. Most CR estimators for that type of applications assume the modes are linear chirps [23], therefore associated with constant CRs. On the contrary, in speech processing, high-resolution speech analysis needs the definition of local CR estimators in particular in applications such as speech transformation and objective voice function assessment for detection of voice disorders [24]. Moreover, for voice signals CR is related to the *jitter* phenomenon [25], [26]. Classically, the local CR estimators are derived by windowing the signal, assuming the latter is locally stationary. Note that, contrary to the local CR estimators used in second and higher order synchrosqueezing transforms, all the just mentioned CR estimators are parametric and therefore non-adaptive. Unfortunately, the local CR estimator used in the definition of the second order synchrosqueezing transform is very unstable in the presence of noise. So, our first concern

N. Laurent and S. Meignen are with the Jean Kuntzmann Laboratory, University Grenoble Alpes and CNRS 5225, Grenoble 38401, France (emails: nils.laurent1@univ-grenoble-alpes.fr, sylvain.meignen@univ-grenoble-alpes.fr). M. A. Colominas is with the Institute for Research and Development in Bioengineering and Bioinformatics (IBB), CONICET, Ruta Prov. 11 Km. 10, Oro Verde, Entre Ríos, Argentina (e-mail: macolominas@conicet.gov.ar). This work was supported in part by the University Grenoble Alpes under IRS Grant "AMUSETE" and the ANR ASCETE project with grant number ANR-19-CE48-0001-01.

here is to find a way to reduce this instability, while preserving the adaptive nature of the local CR estimation. Once this is done, we propose two applications of our improved local CR estimator, one on mode reconstruction based on linear chirp approximation [27], and another one on the analysis of frequency variations in voice signals.

After having briefly recalled, in Section II, some useful definitions regarding STFT, the *multicomponent signals* (MCSs) we are dealing with, and the local CR estimator used in second order synchrosqueezing transform, we illustrate the oscillatory behavior of the latter in the presence of noise in Section III. Based on the definition of a simpler local CR estimator, behaving similarly to the original one in the presence of noise, we characterize more precisely the oscillations of the original CR estimator, and finally find a way to filter them out to obtain a much more accurate CR estimator. In the final section of the paper, we investigate the two different applications mentioned in the previous paragraph.

II. Definitions and Notations

A. Short-Time Fourier Transform

In this section, we introduce a series of definitions that will be used throughout the paper. Considering a signal $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and a real window $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, where $L^p(\mathbb{R})$ stands for the functions h such that $\int_{\mathbb{R}} |h(t)|^p dt < +\infty$, the (modified) short-time Fourier transform (STFT) is defined as:

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

The original signal f can then be retrieved through the following synthesis formula, provided $g(0) \neq 0$:

$$f(t) = \frac{1}{g(0)} \int_{\mathbb{R}} V_f^g(t,\eta) d\eta. \tag{2}$$

The definition of the STFT given in (1) can be easily extended to tempered distributions [28]. Indeed, when one considers an idealized zero-mean complex white Gaussian noise $n(t)$ with covariance matrix $\sigma_n^2 I$, the integral $\int_{\mathbb{R}} n(\tau)\psi(\tau - t)d\tau$, for a function $\psi \in L^2(\mathbb{R})$ should be interpreted as an Itô integral (i.e. as $\int \psi(\tau - t) dB_\tau$, the integration with respect to 1-parameter Brownian motion B_τ). Therefore, such an integration results in a complex Gaussian random variable with zero-mean and covariance matrix $\sigma_n^2 \|\psi\|^2 I$ [29], with $\|\psi\|^2 = \int_{\mathbb{R}} \psi^2(t) dt$ being the squared L^2-norm of ψ, and $V_n^g(t,\eta)$ (the STFT of complex white Gaussian noise $n(t)$) is well defined.

In the sequel, we will use multicomponent signals (MCSs) defined as the superimposition of AM/FM components, namely:

$$f(t) = \sum_{p=1}^{P} f_p(t) \text{ with } f_p(t) = A_p(t) e^{2i\pi\phi_p(t)}, \tag{3}$$

in which $A_p(t)$ and $\phi_p(t)$ correspond respectively to the instantaneous amplitude (IA) and phase (IP) of the p^{th} mode. We assume that $A_p(t) > 0$, $\phi_p'(t) > 0$ and $\phi_{p+1}'(t) > \phi_p'(t)$, where $\phi_p'(t)$ denotes the instantaneous frequency (IF) of f_p at time t.

B. Local Chirp Rate Estimator Used in Second Order Synchrosqueezing Transforms

The definition of the local CR estimator used in the second order synchrosqueezing transform uses two *complex reassignment operators* $\tilde{\omega}_f(t,\eta) = \frac{\partial_t V_f^g(t,\eta)}{2i\pi V_f^g(t,\eta)}$ and $\tilde{t}_f(t,\eta) = t - \frac{\partial_\eta V_f^g(t,\eta)}{2i\pi V_f^g(t,\eta)}$. It can be defined following [21] as (we omit (t,η) for the sake of simplicity):

$$\begin{aligned} \widehat{q}_f = & \quad \Re\left\{\frac{\partial_\eta \tilde{\omega}_f}{\partial_\eta \tilde{t}_f}\right\} = \Re\left\{\frac{\partial_\eta\left(\frac{\partial_t V_f^g}{V_f^g}\right)}{2i\pi - \partial_\eta\left(\frac{\partial_\eta V_f^g}{V_f^g}\right)}\right\} \\ = & -\frac{1}{2\pi}\Im\left\{\frac{\left(V_f^g\right)^2 - V_f^{g'} V_f^{tg} + V_f^g V_f^{tg'}}{V_f^g V_f^{t^2 g} - \left(V_f^{tg}\right)^2}\right\}, \end{aligned} \tag{4}$$

where $\Re$ (resp. $\Im$) denotes the real (resp. imaginary) part. Note that $\widehat{q}_f$ is exact if f is a Gaussian modulated linear chirp.

When the window g is the Gaussian function $g(t) = e^{-\pi\frac{t^2}{\sigma^2}}$, we have $V_f^{g'} = -\frac{2\pi}{\sigma^2} V_f^{tg}$ and thus $\frac{-V_f^{g'} V_f^{tg} + V_f^g V_f^{tg'}}{V_f^g V_f^{t^2 g} - \left(V_f^{tg}\right)^2}$ is a real number, and thus we obtain the simpler expression:

$$\widehat{q}_f = -\frac{1}{2\pi}\Im\left\{\frac{(V_f^g)^2}{V_f^g V_f^{t^2 g} - (V_f^{tg})^2}\right\}. \tag{5}$$

We shall remark here that, in the seminal paper introducing the second order synchrosqueezing transform [30], the local CR estimate was defined as $\Re\{\frac{\partial_t \tilde{\omega}_f}{\partial_t \tilde{t}_f}\}$, which is equal to $\widehat{q}_f$ defined in (5) when g is a Gaussian window [31].

III. Analysis of Local Chirp Rate Estimate

In this section, our goal is to investigate the behavior of $\widehat{q}_f$, defined in (5), in the vicinity of the STFT ridges associated with the modes of an MCS. After having noticed that this CR estimator is very oscillatory in the presence of noise, we are going to define a simplified local CR estimator which behaves similarly to the initial CR estimator in the presence of noise on STFT ridges. Such a definition will enable us to analyze the nature of the oscillations in the initial CR estimator and then to filter them out.

A. Definition of a Simplified Local CR Estimator

Our motivation for the definition of a simplified version of $\widehat{q}_f$ is based on the fact that, if one adds a complex white Gaussian noise n to f, then $\widehat{q}_{f+n}$ is very oscillatory on the STFT ridges of the noisy signal, as illustrated in Fig. 1 second row. Indeed, on the first row of that figure, we display the STFT moduli of three different mono-component signals, along with the STFT ridge associated with the mode in each case. On the second row of that figure, we display a zoomed-in version in time of $\widehat{q}_{f+n}$ on STFT ridges and the true CR, namely $\phi''(t)$, and we notice that, for the three signals, the former oscillates a lot around the expected value. Our goal is thus to define an approximation of $\widehat{q}_{f+n}$ on STFT ridges that would help us to understand the nature of these oscillations.

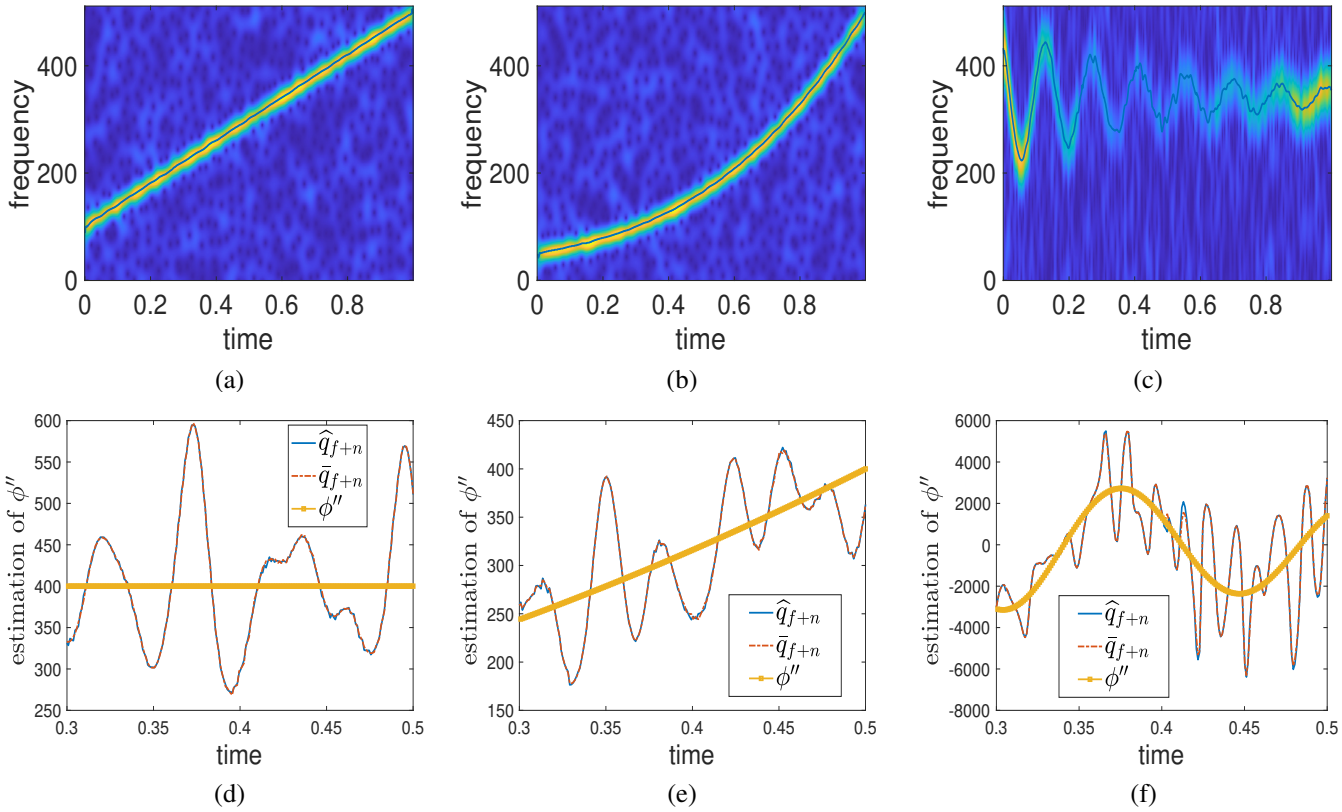


Fig. 1: (a): STFT modulus of a noisy linear chirp with the STFT ridge superimposed; (b): same as (a) but for a noisy chirp with fourth order polynomial phase; (c): same as (a) but for a noisy mode with oscillating phase; (d): CR estimator $\widehat{q}_{f+n}$ or $\bar{q}_{f+n}$ estimated on the STFT ridge 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.

For that purpose, let us consider that $f(t)$ is a linear chirp $Ae^{2i\pi\phi(t)}$, with $\phi''(t) = C$. For such a signal, since one has [19]

$$V_f^g(t,\eta) = V_f^g(t,\phi'(t))e^{\frac{-\pi\sigma^2(1+i\phi''(t)\sigma^2)}{1+(\phi''(t)\sigma^2)^2}(\eta-\phi'(t))^2}, \qquad (6)$$

one obtains that $V_f^{tg}(t,\phi'(t)) = \frac{i}{2\pi}\partial_\eta V_f^g(t,\phi'(t)) = 0$. So, one can reasonably assume that, when (t,η) is in the vicinity of the STFT ridge, namely the TF curve $(t,\phi'(t))$, $|V_f^{tg}(t,\eta)|$ is small compared with both $|V_f^g(t,\eta)|$ and $|V_f^{t^2g}(t,\eta)|$. Doing a first order Taylor expansion of $\widehat{q}_f(t,\eta)$ with respect to the variable $\frac{(V_f^{tg}(t,\eta))^2}{V_f^g(t,\eta)V_f^{t^2g}(t,\eta)}$ leads to the following approximation of $\widehat{q}_f$ (we omit (t,η) for the sake of simplicity):

$$\widehat{q}_f = -\frac{1}{2\pi}\Im\left\{\frac{V_f^g}{V_f^{t^2g}}\frac{1}{1-\frac{(V_f^{tg})^2}{V_f^gV_f^{t^2g}}}\right\} \approx -\frac{1}{2\pi}\Im\left\{\frac{V_f^g}{V_f^{t^2g}}\right\}. \qquad (7)$$

In what follows, we will denote by $\bar{q}_f$ the approximation of $\widehat{q}_f$ associated with equation (7). Note that, because V_f^{tg} is null on STFT ridge and as $\widehat{q}_f(t,\phi'(t)) = \phi''(t)$, one also has $\bar{q}_f(t,\phi'(t)) = \phi''(t)$.

Now, adding a complex white Gaussian noise n to f, one can consider, in the vicinity of the ridge of the noisy STFT, the local CR estimator:

$$\bar{q}_{f+n} = -\frac{1}{2\pi}\Im\left\{\frac{V_f^g + V_n^g}{V_f^{t^2g} + V_n^{t^2g}}\right\}, \qquad (8)$$

which appears to be very close numerically to $\widehat{q}_{f+n}$, on the STFT ridges of noisy signals and even at a high noise level (see the second row of Fig. 1 for illustrations, where the input SNR is 10 dB). To clarify this aspect, we plot, in Fig. 2, $\frac{\|\widehat{q}_{f+n}-\bar{q}_{f+n}\|_2}{\|\widehat{q}_{f+n}\|_2}$ for the three signals corresponding to the first row of Fig. 1, and observe that this ratio is very low at a low noise level for the three signals and remains so at high noise level for the first two signals, while the ratio increases slightly for the last signal but only at a high noise level. To investigate the oscillations of $\widehat{q}_{f+n}$, we study, in the next section, $\bar{q}_{f+n}$ which is much more amenable to mathematical study than $\widehat{q}_{f+n}$.

B. Study of the Simplified CR Estimate

We here study more in detail the local CR estimator $\bar{q}_{f+n}$. For that purpose, we first compute the estimation bias and then explain which terms are responsible for the oscillations in the estimation. To characterize the bias, let us first set $Z(t,\eta) = \frac{V_{f+n}^g(t,\eta)}{V_{f+n}^{t^2g}(t,\eta)}$ and then study $-\frac{1}{2\pi}\Im\{Z(t,\eta)\}$, for which we have

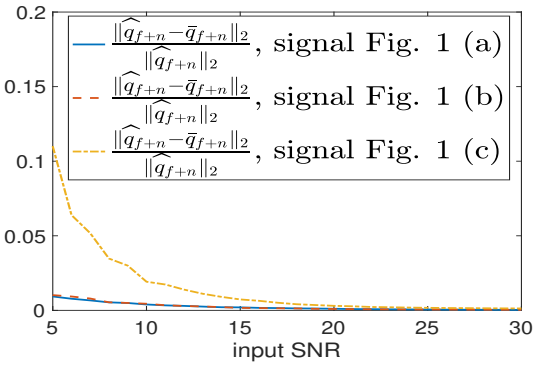


Fig. 2: comparison of $\widehat{q}_{f+n}$, with $\bar{q}_{f+n}$ on the STFT ridge of the noisy signal, for the three signals of Fig. 1.

the following proposition (omitting $(t,\phi'(t))$ for the sake of simplicity):

Proposition III.1. *At a STFT ridge point $(t,\phi'(t))$ of a linear chirp one has:*

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

where σ_n^2 is the variance of the noise.

The proof is available in Appendix A. From (6) and from the fact that $V_f^{tg}(t,\eta) = \frac{i}{2\pi}\partial_\eta V_f^g(t,\eta)$, we get that for a linear chirp $f(t) = Ae^{2i\pi\phi(t)}$, one has [31]

$$V_f^{t^kg}(t,\eta) = \left(\frac{i}{2\pi}\right)^k \partial_\eta^k V_f^g(t,\eta) = \frac{-\sigma^2(1+i\phi''(t)\sigma^2)}{1+(\phi''(t)\sigma^2)^2} \left[i(\eta-\phi'(t))V_f^{t^{k-1}g}(t,\eta) - \frac{k-1}{2\pi}V_f^{t^{k-2}g}(t,\eta)\right], \tag{9}$$

and thus,

$$V_f^{t^kg}(t,\phi'(t)) = \frac{\sigma^2(1+i\phi''(t)\sigma^2)}{1+(\phi''(t)\sigma^2)^2}\frac{k-1}{2\pi}V_f^{t^{k-2}g}(t,\phi'(t)), \tag{10}$$

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

$$|V_f^{t^2g}| = \frac{\sigma^2\left|V_f^g\right|}{2\pi\sqrt{1+(\phi''(t)\sigma^2)^2}} = \frac{A\sigma^2}{2\pi(1+(\phi''(t)\sigma^2)^2)^{3/4}}. \tag{11}$$

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

$$\frac{|V_f^{t^2g}|^2}{\|t^2g\|^2} = \frac{A^2 4\sqrt{2}}{3\sigma(1+(\phi''(t)\sigma^2)^2)^{3/2}}. \tag{12}$$

As from Proposition III.1 when $\sigma_n \ll \frac{|V_f^{t^2g}|}{\|t^2g\|}$ the estimator is only slightly biased on the STFT ridge, (12) tells us that the bias increases when the modulation of the mode increases. Nevertheless, for the linear chirp of Fig 1. (a) and for an input SNR of 0 dB, averaging over many realizations of the noise, we get that $\bar{q}_{f+n}$ is very slightly biased, and this small estimation bias is not responsible for the oscillations put forward in the previous section.

Assuming that $\mathbb{E}\left\{\frac{|V_n^{t^2g}|}{|V_f^{t^2g}|}\right\} = \frac{\sigma_n\|t^2g\|}{|V_f^{t^2g}|}$ is small in the vicinity of the STFT ridges of noisy signals, we do a first order Taylor expansion of $\bar{q}_{f+n}$ with respect to the variable $\frac{V_n^{t^2g}}{V_f^{t^2g}}$ at these TF locations, to get:

$$\begin{aligned}\bar{q}_{f+n} &= -\frac{1}{2\pi}\Im\left\{\frac{V_f^g+V_n^g}{V_f^{t^2g}+V_n^{t^2g}}\right\}\\ &= -\frac{1}{2\pi}\Im\left\{\frac{V_f^g}{V_f^{t^2g}}\frac{1}{1+\frac{V_n^{t^2g}}{V_f^{t^2g}}}+\frac{V_n^g}{V_f^{t^2g}}\frac{1}{1+\frac{V_n^{t^2g}}{V_f^{t^2g}}}\right\}\\ &\approx -\frac{1}{2\pi}\Im\left\{\frac{V_f^g}{V_f^{t^2g}}(1-\frac{V_n^{t^2g}}{V_f^{t^2g}})+\frac{V_n^g}{V_f^{t^2g}}(1-\frac{V_n^{t^2g}}{V_f^{t^2g}})\right\}\\ &\approx \bar{q}_f+\frac{1}{2\pi}\Im\left\{\frac{V_f^gV_n^{t^2g}}{(V_f^{t^2g})^2}-\frac{V_n^g}{V_f^{t^2g}}\right\},\end{aligned} \tag{13}$$

the last approximation being obtained by assuming $\Im\left\{\frac{V_n^gV_n^{t^2g}}{(V_f^{t^2g})^2}\right\}$ is negligible compared with the other noise contributions. In what follows, we will denote by $\widetilde{q}_{f+n}$ the approximation of $\bar{q}_{f+n}$ given by (13). We numerically notice that $\widetilde{q}_{f+n}$ is still very close to $\widehat{q}_{f+n}$ when evaluated on the STFT ridges of noisy linear chirps whatever the noise level as shown in Fig. 3, the difference being only significant when the mode has a fast oscillating phase and at low SNRs.

Thus, as $\bar{q}_f$ equals ϕ'' on the STFT ridge of a linear chirp, the oscillations for more general signals are mainly due to the term $\frac{1}{2\pi}\Im\left\{\frac{V_f^gV_n^{t^2g}}{(V_f^{t^2g})^2}-\frac{V_n^g}{V_f^{t^2g}}\right\}$, which we now analyze more in detail. As on the STFT ridge of a linear chirp, from (10), one has $\frac{V_f^g(t,\phi'(t))}{V_f^{t^2g}(t,\phi'(t))} = \frac{2\pi}{\sigma^2}(1-i\phi''(t)\sigma^2)$, thus one may write (omitting $(t,\phi'(t))$ for the sake of simplicity):

$$\begin{aligned}\Im\left\{\frac{V_f^gV_n^{t^2g}}{(V_f^{t^2g})^2}-\frac{V_n^g}{V_f^{t^2g}}\right\} &= \Im\left\{\frac{V_f^g}{V_f^{t^2g}}\left(\frac{V_n^{t^2g}}{V_f^{t^2g}}-\frac{V_n^g}{V_f^g}\right)\right\}\\ &= \Im\left\{\frac{V_f^g}{V_f^{t^2g}}\right\}\Re\left\{\frac{V_n^{t^2g}}{V_f^{t^2g}}-\frac{V_n^g}{V_f^g}\right\}\\ &\quad+\Re\left\{\frac{V_f^g}{V_f^{t^2g}}\right\}\Im\left\{\frac{V_n^{t^2g}}{V_f^{t^2g}}-\frac{V_n^g}{V_f^g}\right\}\\ &= -2\pi\phi''(t)\Re\left\{\frac{V_n^{t^2g}}{V_f^{t^2g}}-\frac{V_n^g}{V_f^g}\right\}+\frac{2\pi}{\sigma^2}\Im\left\{\frac{V_n^{t^2g}}{V_f^{t^2g}}-\frac{V_n^g}{V_f^g}\right\},\end{aligned} \tag{14}$$

meaning the oscillations are related to $G(t) := \frac{V_n^{t^2g}}{V_f^{t^2g}}-\frac{V_n^g}{V_f^g}$, which we analyze more carefully.

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:

$$G(t) = \frac{2\pi}{\sigma^2}\frac{V_n^{t^2g}}{V_f^g}-\frac{V_n^g}{V_f^g} = \left(\frac{2\pi}{\sigma^2}V_n^{t^2g}-V_n^g\right)e^{-2i\pi\eta_0 t}, \tag{15}$$

to get the following:

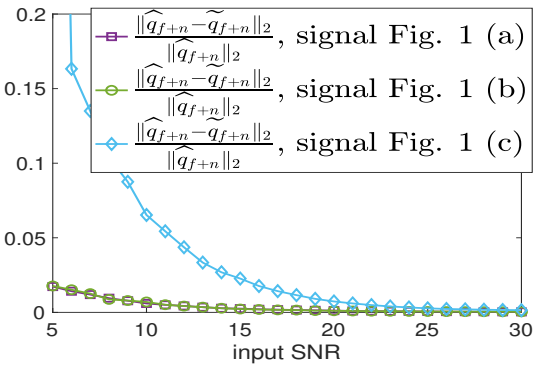


Fig. 3: comparison of $\widehat{q}_{f+n}$ with $\widetilde{q}_{f+n}$ on the STFT ridge of the noisy signal, for the three signals of Fig. 1.

Proposition III.2. *The power spectral density of $G(t)$ is*

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

The proof is available in Appendix B. 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 (10),

$$\frac{V_f^g}{V_f^{t^2 g}} = \frac{2\pi(1+(b\sigma^2)^2)}{\sigma^2(1+ib\sigma^2)} = \frac{2\pi}{\sigma^2}(1-ib\sigma^2), \tag{16}$$

and thus:

$$G(t) = \frac{\frac{2\pi}{\sigma^2}(1-ib\sigma^2)V_n^{t^2 g} - V_n^g}{V_f^g}, \tag{17}$$

leading to the following:

Proposition III.3. *The power spectral density of $G(t)$ is*

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

The proof is available in Appendix C.

C. Low-Pass Filtering Local CR Estimator

In this section, we investigate how to practically filter out the oscillations in the local CR estimator $\widehat{q}_{f+n}$ using a low-pass filter, designed using the power spectral densities of the function G derived in the previous section.

For that purpose, we remark first that, when ϕ'' is null, the square root of P_G, denoted by H_G in what follows, passes through a maximum at $\eta_m = \frac{1}{\sqrt{\pi}\sigma}$, and $H_G(\eta_m) = 2\sigma\sigma_n e^{-1}$. Then, one remarks that the frequency $\eta_{c,0}$ at which $H_G(\eta_{c,0}) = cH_G(\eta_m)$ (with $c<1$ representing a fraction of the maximum value) corresponds to

$$\begin{aligned} \sigma_n\sigma^3 2\pi\eta_{c,0}^2 e^{-\pi\sigma^2\eta_{c,0}^2} &= 2\sigma\sigma_n c e^{-1} \\ \Leftrightarrow \eta_{c,0} &= \frac{\sqrt{-\mathcal{W}(-ce^{-1})}}{\sigma\sqrt{\pi}}, \end{aligned} \tag{18}$$

with $\mathcal{W}(\cdot)$ the Lambert W function [32]. So, to filter out the oscillations, we propose to low-pass filter $\widehat{q}_{f+n}$ by multiplying its frequency spectrum with $\mathbf{1}_{[0,\eta_{c,0}]}$.

When $\phi''=b$, H_G attains its maximum at $\eta_m = \frac{\sqrt{1+b^2\sigma^4}}{\sigma\sqrt{\pi}}$, with $H_G(\eta_m) = 2\sigma_n\sigma e^{-1}$, and thus the frequency $\eta_{c,b}$ at which $H_G(\eta_{c,b}) = cH_G(\eta_m)$ corresponds to

$$\begin{aligned} \frac{\sigma_n\sigma^3 2\pi\eta_{c,b}^2}{(1+b^2\sigma^4)} e^{-\frac{\pi\sigma^2\eta_{c,b}^2}{1+b^2\sigma^4}} &= 2c\sigma_n\sigma e^{-1} \\ \Leftrightarrow \eta_{c,b} &= (1+\sigma^4 b^2)^{1/2}\frac{\sqrt{-\mathcal{W}(-ce^{-1})}}{\sigma\sqrt{\pi}}, \end{aligned} \tag{19}$$

which generalizes the previous result. From this last equation, we observe that the chirp rate $\phi''=b$ shifts the cut-off frequency to the right. However, since in practice we do not have access to b, we will consider filtering the oscillations in $\widehat{q}_{f+n}$ still by multiplying its frequency spectrum with $\mathbf{1}_{[0,\eta_{c,0}]}$. This will prove enough for most of the signals encountered in practice. For example, if the signal is a linear chirp, then $\widehat{q}_f$ is a constant function (with frequency content only at zero frequency), and the low-pass filtering process will keep all the signal-related information, repelling most of the noise. We show that this filtering process is good enough also for other types of signals. From now on, we consider that $F(\widehat{q}_{f+n})$ is the filtered version of $\widehat{q}_{f+n}$ using the cut-off frequency $\eta_{c,0}$.

IV. Results

The purpose of this section is three-fold. The first one is to investigate the quality of the filtering process proposed in Section III-C to see how $F(\widehat{q}_{f+n})$ improves the original CR estimator $\widehat{q}_{f+n}$ in various situations. Note that to improve CR estimation can be profitable to CR-based ridge detection [33], to mode separation in case of interference [34] and to mode retrieval based on linear chirp approximation [27], to name a few. Therefore, a second objective of this section is to show in what way the CR estimator $F(\widehat{q}_{f+n})$ improves the mode retrieval technique based on linear chirp approximation proposed in [27]. Finally, we will see how a better CR estimator could be profitably used in the study of voice signals.

A. Evaluating the Filtering Process

We would like to assess the improvement brought by the low-pass filtering process to CR estimation, namely to compare $F(\widehat{q}_{f+n})$ to $\widehat{q}_{f+n}$, on the STFT ridges of a wide range of noisy signals. To do so, we compare $F(\widehat{q}_{f+n})$ and $\widehat{q}_{f+n}$ to the actual $\phi''(t)$ for the three signals of Fig. 1, when the input SNR varies. The quality of the estimation is measured in terms of the output SNR, namely:

$$SNR_{out}(\varphi,\phi'') = 10\log_{10}\left(\frac{\|\phi''\|^2}{\|\varphi-\phi''\|^2}\right), \tag{20}$$

where φ is either $\widehat{q}_{f+n}$ or $F(\widehat{q}_{f+n})$ both evaluated on the STFT ridges of noisy signals. The results depicted in Fig. 4 first row, for the three signals of Fig. 1, show that filtering $\widehat{q}_{f+n}$ brings an important estimation improvement. These figures also tell us that the variance of the filtered estimate remains low, that the worst filtered estimates not considered as outliers are always better than the best estimates without filtering not considered as outliers, and finally that there are very few outliers in both estimation processes (the outliers correspond to red stars on the graphs of the first row of Fig. 4). The only notable exception is for the signal associated with of Fig. 1 (c) for which, at 0 dB, ridge detection may fail which makes

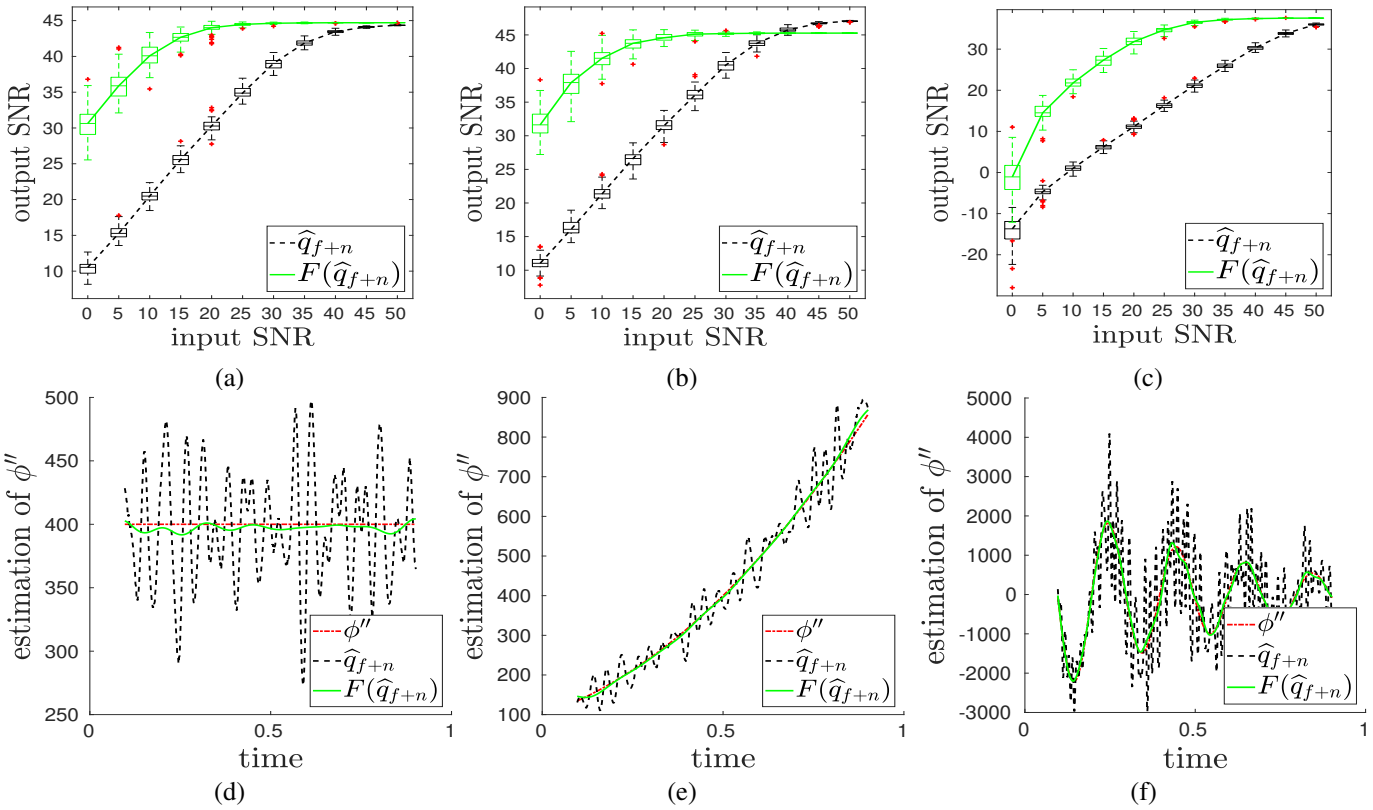


Fig. 4: (a): boxplot corresponding to the output SNR associated with CR estimators $\widehat{q}_{f+n}$ or $F(\widehat{q}_{f+n})$ evaluated on the STFT ridges of the noisy signal corresponding to Fig. 1 (a) over 100 realizations of the noise when the input SNR varies; (b): same as (a) but for the noisy signal corresponding to Fig. 1 (b); (c): same as (a) but for the noisy signal corresponding to Fig. 1 (c); (d): $\widehat{q}_{f+n}$ and $F(\widehat{q}_{f+n})$ computed for a noisy version of the signal of Fig. 1 (a), along with the ground truth $\phi''(t)$ (input SNR = 10 dB); (e): same as (d) but for the signal of Fig. 1 (b); (f): same as (d) but for the signal of Fig. 1 (c).

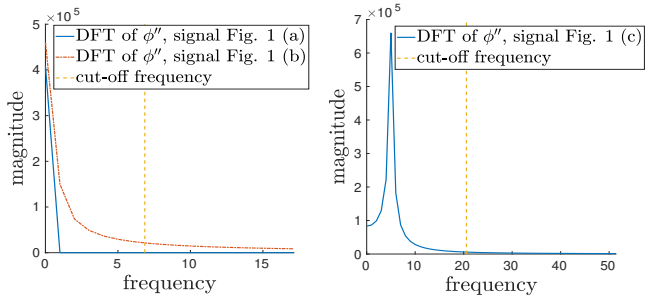


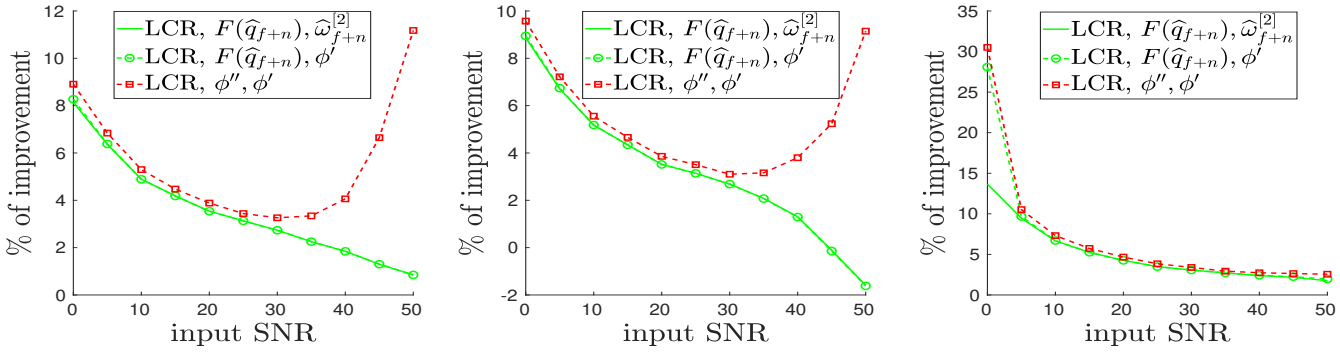
Fig. 5: (a): magnitude of the DFT of ϕ'' for the signals of Fig. 1 (a) and Fig. 1 (b); (b): same as (a) but for the signal of Fig. 1 (c).

CR estimation results hard to interpret. We may also remark that the results for the signal of Fig. 1 (a) and (b), displayed in Fig. 4 (a) and (b), are pretty much the same which means that to filter out the oscillations using the cut-off frequency $\eta_{c,0}$ is efficient even when the mode is frequency modulated.

On the second row of Fig. 4, we plot illustrations of $F(\widehat{q}_{f+n})$ and $\widehat{q}_{f+n}$ computed on noisy versions of the signals of Fig. 1 along with the ground truth ϕ'' (the input SNR equals 10 dB in each case). We notice that the proposed procedure succeeds in filtering out most of the oscillations of $\widehat{q}_{f+n}$. However, because those three signals contain different types of frequency modulations, we now check that the choice of the cut-off frequency only removes the oscillations associated with the noise. In the simulations, we considered $c = 1/\sqrt{10}$ in the definition of $\eta_{c,0}$, meaning that we take the cut-off frequency at the 10% of the maximum energy. By considering the multiplication of the frequency spectrum of $\widehat{q}_{f+n}$ by $\mathbf{1}_{[0,\eta_{c,0}]}$, one removes all the more oscillations that the signal is modulated. However, we should also check that the proposed threshold does not damage too much the frequencies associated with ϕ''. In this regard, on Fig. 5, we display the *discrete Fourier transform* (DFT) of ϕ'' for the three signals of Fig. 1, along with the cut-off frequency computed with $c = 1/\sqrt{10}$. In each case, the frequencies corresponding to the signal are very well-preserved, which tells us that this value for c is suitable. Therefore, in the following simulations, we will keep this value for parameter c.

B. Improving Linear Chirp Based Mode Retrieval

In this section, we investigate how to use the filtered CR estimator defined in Section III-C to improve *linear chirp based mode retrieval* (LCR) technique introduced in [27]. In a nutshell, this technique assumes a local linear chirp approximation for a mode, thus its STFT can be approximated



(a) (b) (c)

Fig. 6: (a): percentage of improvement for the LCR technique computed using (28) for the signal of Fig. 1 (a) where the estimates for (ϕ'', ϕ') are $(F(\widehat{q}_{f+n}(t, R_{f+n}(t))), \omega^{[2]}_{f+n}(t, R_{f+n}(t)))$, $(F(\widehat{q}_{f+n}(t, R_{f+n}(t))), \phi')$, or (ϕ'', ϕ');(b): same as (a) but for the signal of Fig. 1 (b) ;(c): same as (a) but for the signal of Fig. 1 (c)

by [19]:

$$V^g_f(t,\eta) \approx V^g_{f,approx}(t,\eta) := A(t) e^{-\pi\sigma^2 \frac{1+iC(t)\sigma^2}{1+(C(t)\sigma^2)^2}(\eta - B(t))^2}, \tag{21}$$

where $A(t)$, $B(t)$ and $C(t)$ are respectively estimates of $V^g_f(t, \phi'(t))$, $\phi'(t)$ and $\phi''(t)$. In that context, mode reconstruction is performed by summing the approximation given by (21) over frequencies to obtain:

$$f(t) \approx \frac{1}{g(0)} \int_{\mathbb{R}} V^g_{f,approx}(t,\eta) d\eta, \tag{22}$$

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 $\widehat{\omega}^{[2]}_f = \Re\left\{\tilde{\omega}^{[2]}_f\right\}$ used in the definition of the second order synchrosqueezing transform [19], with:

$$\tilde{\omega}^{[2]}_f = \begin{cases} \tilde{\omega}_f + \frac{\partial_\eta \tilde{\omega}_f}{\partial_\eta \tilde{t}_f} \times (t - \tilde{t}_f) \text{ if } \partial_\eta \tilde{t}_f \neq 0 \\ \tilde{\omega}_f \text{ otherwise,} \end{cases} \tag{23}$$

where $\tilde{\omega}_f$ and $\tilde{t}_f$ were introduced in Section II-B. More precisely, denoting by $R_f(t)$ the STFT ridge corresponding to f, $B(t)$ is set to $\widehat{\omega}^{[2]}_f(t, R_f(t))$, $C(t)$ to $\widehat{q}_f(t, R_f(t))$, and finally $A(t)$ to $V^g_f(t, \widehat{\omega}^{[2]}_f(t, R_f(t)))$. When one considers the noisy signal $f + n$ with n a complex Gaussian white noise, similar estimates are derived replacing f by $f + n$ 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 $F(\widehat{q}_{f+n}(t, R_{f+n}(t)))$ instead of $\widehat{q}_{f+n}(t, R_{f+n}(t))$ 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 $F(\widehat{q}_{f+n}(t, R_{f+n}(t)))$ or to $\phi''(t)$. To clarify this, as LCR can be viewed as a function of 3 parameters, we define:

$$LCR_{A,B,C}(t,\eta) := A(t) e^{-\pi\sigma^2 \frac{1+iC(t)\sigma^2}{1+(C(t)\sigma^2)^2}(\eta - B(t))^2}. \tag{24}$$

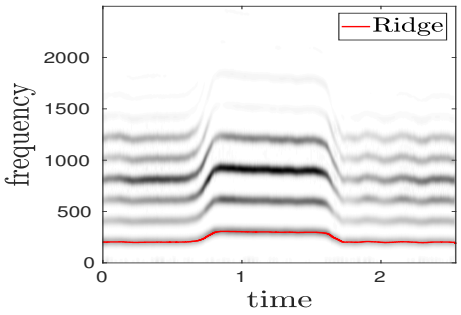


Fig. 7: ridge corresponding to the fundamental frequency superimposed on the STFT of the recording 62 taken from [35].

In that framework, the original LCR technique corresponds to

$$\begin{aligned} (A,B,C) &= (A_0, B_0, C_0) \\ &:= (V^g_{f+n}(t, \widehat{\omega}^{[2]}_{f+n}(t, R_{f+n}(t))), \\ &\quad \widehat{\omega}^{[2]}_{f+n}(t, R_{f+n}(t)), \widehat{q}_{f+n}(t, R_{f+n}(t))), \end{aligned} \tag{25}$$

the proposed technique to:

$$(A,B,C) = (A_0, B_0, F(\widehat{q}_{f+n}(t, R_{f+n}(t)))), \tag{26}$$

while the other two references to

$$\begin{aligned} (A,B,C) &= (A_0, \phi'(t), F(\widehat{q}_{f+n}(t, R_{f+n}(t)))) \\ (A,B,C) &= (A_0, \phi'(t), \phi''(t)). \end{aligned} \tag{27}$$

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 (20), we compute the percentage of improvement brought by using a set (A, B, C) different from (A_0, B_0, C_0) as follows:

$$\frac{SNR_{out}(f_{(A,B,C)}, f) - SNR_{out}(f_{(A_0,B_0,C_0)}, f)}{SNR_{out}(f_{(A_0,B_0,C_0)}, f)} \times 100. \tag{28}$$

The results are depicted in Fig. 6 for the three signals of Fig. 1 and when (A, B, C) are either defined by (26) or (27). First, these clearly show the improvement brought by using

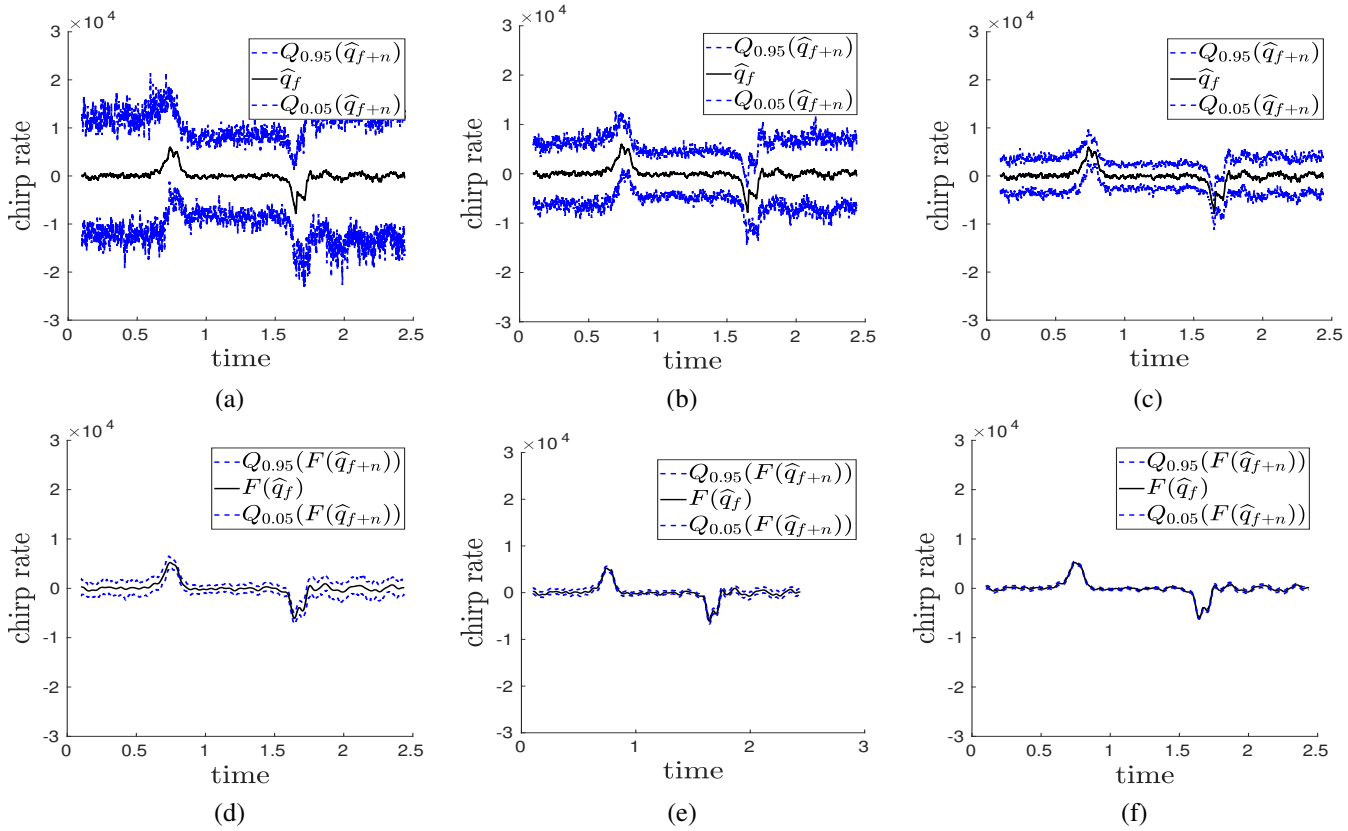


Fig. 8: (a): for 100 realizations of the noise, the range corresponding to the 5% and 95% quantiles of $\widehat{q}_{f+n}$ evaluated on the STFT ridge corresponding to the fundamental frequency of the noisy signal whose STFT modulus is displayed in Fig. 7, $\widehat{q}_f$ on the ridge corresponding to 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 $F(\widehat{q}_{f+n})$ instead of $\widehat{q}_{f+n}$, $F(\widehat{q}_f)$ is also superimposed ; (e): Same as (d) but with input SNR$= 10$ dB; (f): same as (d) but with input SNR$= 15$ dB.

$F(\widehat{q}_{f+n}(t, R_{f+n}(t)))$ instead of $\widehat{q}_{f+n}(t, R_{f+n}(t))$. 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 ϕ'' 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 $(F(\widehat{q}_{f+n}(t, R_{f+n}(t))), \widehat{\omega}^{[2]}_{f+n}(t, R_{f+n}(t, R_{f+n}(t))))$ for (ϕ'', ϕ') 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 $V^g_f(t, \phi'(t))$ than $A_0(t)$, but this is beyond the scope of the present paper.

C. Application to Voice Signals

Time-frequency analysis techniques have been used on real voice signals, and in particular, some rely on a CR estimator as in [36] for mode reconstruction. Moreover, CR estimators have been used to redefine the so-called *jitter* parameter, obtaining results of the state of the art [26]. In this section, we investigate the robustness to noise of CR estimators $\widehat{q}_f$ and $F(\widehat{q}_f)$ on the voice signal of a 22-year-old female taken from the recording 62 of the Saarbruken Voice Database [35]. 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 in the pitch of the signal. The modulus of this STFT is depicted in Fig. 7 along with the ridge corresponding to the fundamental frequency of this signal.

To test the robustness to noise of the CR estimators mentioned above, we study how the same additive noise n affects the original CR estimation made on the signal without noise. Namely, we compare the stability of $\widehat{q}_{f+n}(t, R_{f+n}(t))$ with that of $F(\widehat{q}_{f+n}(t, R_{f+n}(t)))$ using $\widehat{q}_f(t, R_f(t))$ and $F(\widehat{q}_f(t, R_f(t)))$ as respective references, and in which $R_f(t)$ and $R_{f+n}(t)$ are the ridges corresponding to the fundamental frequency of the noise-free and noisy signal respectively. To carry out this study, we compute the 5% and 95% quantiles over 100 realizations of noise for three different input SNRs: 5, 10 and 15 dB. The results depicted in Fig. 8 show that $F(\widehat{q}_{f+n}(t, R_{f+n}(t)))$ is significantly more stable than

$\widehat{q}_{f+n}(t, R_{f+n}(t))$ since both quantiles are much closer to the reference. Furthermore, it is important to note that the filtering process preserved the variation of the pitch present at 0.75s and 1.65s.

V. Conclusion

In this paper, our goal was to improve the chirp rate estimator used in the second order synchrosqueezing transform, which appeared very oscillatory in the presence of noise. For that purpose, we made reasonable assumptions that allowed us to define a simplified chirp rate estimator which enabled us to understand where these oscillations came from, and then to filter them out to obtain a more accurate chirp rate estimator. This filtering procedure turned out to be efficient even at low input SNR and for various types of signals. Secondly, we used this new chirp rate estimator to improve mode reconstruction based on linear chirp approximation. Finally, we proposed to use the new chirp rate estimator on a voice signal taken from the Saarbruken Voice Database and showed that the filtering process used to design the new estimator resulted in a stable estimation, making easier the following of pitch variations. A potential perspective for this work would be to further extend the filtering procedure to improve the estimation of the short-time Fourier transform on the ridges associated with the modes of noisy multicomponent signals to improve mode reconstruction.

Appendix A

In what follows, we omit (t, η) in the definition of the STFTs. Let us first remark that:

$$\mathbb{E}\{V^g_{f+n}\} = V^g_f, \quad \mathbb{E}\{V^{t^2g}_{f+n}\} = V^{t^2g}_f$$

since V^g_n and $V^{t^2g}_n$ are both with zero mean. Let us denote by $\mathbf{m} = (V^g_f, V^{t^2g}_f)^T$. Then denoting by $\mathbf{v} = (V^g_{f+n}, V^{t^2g}_{f+n})^T$, and σ_n^2 the variance of the noise (the real and imaginary part of the noise are assumed to be decorrelated), we have:

$$\begin{aligned}\mathbb{E}\{(\mathbf{v}-\mathbf{m})(\mathbf{v}-\mathbf{m})^H\} &= \begin{bmatrix} \mathbb{E}\{|V^g_n|^2\} & \mathbb{E}\{V^g_n (V^{t^2g}_n)^*\} \\ \mathbb{E}\{(V^g_n)^* V^{t^2g}_n\} & \mathbb{E}\{|V^{t^2g}_n|^2\} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_n^2\|g\|^2 & \sigma_n^2\|tg\|^2 \\ \sigma_n^2\|tg\|^2 & \sigma_n^2\|t^2g\|^2 \end{bmatrix}.\end{aligned}$$

Based on the results in [37, Eq. (44)], denoting $Z = \frac{V^g_{f+n}}{V^{t^2g}_{f+n}}$, we have that:

$$\mathbb{E}\{Z\} = \frac{V^g_f}{V^{t^2g}_f} + \left(\frac{\|tg\|^2}{\|t^2g\|^2} - \frac{V^g_f}{V^{t^2g}_f}\right) e^{-\frac{|V^{t^2g}_f|^2}{\sigma_n^2\|t^2g\|^2}},$$

so that one gets:

$$\mathbb{E}\left\{-\frac{1}{2\pi}\Im\{Z\}\right\} = -\frac{1}{2\pi}\Im\left\{\frac{V^g_f}{V^{t^2g}_f}\right\}(1 - e^{-\frac{|V^{t^2g}_f|^2}{\sigma_n^2\|t^2g\|^2}}).$$

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

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

Appendix B

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

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

so we get that:

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

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

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

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:

$$\begin{aligned}\widehat{\tfrac{\sigma^4}{4\pi^2}\sigma_n^2 g''*g''(-x)}(\eta) &= \sigma_n^2\frac{\sigma^4}{4\pi^2}\widehat{g''*g''}(\eta)^* = \sigma_n^2\frac{\sigma^4}{4\pi^2}\widehat{g''}(\eta)^2 \\ &= \sigma_n^2\sigma^4 4\pi^2\eta^4\widehat{g}(\eta)^2 \\ &= \sigma_n^2\sigma^6 4\pi^2\eta^4 e^{-2\pi\sigma^2\eta^2}.\end{aligned}$$

Appendix C

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

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

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

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

so that one may rewrite G as:

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

Then the auto-correlation function of G reads:

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

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

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

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:

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

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

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

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:

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

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

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

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

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

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

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

REFERENCES

[1] L. Cohen, *Time-frequency Analysis: Theory and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[2] P. Flandrin, *Time-frequency/time-scale analysis*. Academic Press, 1998, vol. 10.

[3] 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.

[4] 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.

[5] 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.

[6] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992.

[7] 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.

[8] 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, November 1946.

[9] 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.

[10] 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, Feb 1999.

[11] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Academic Press, 2008.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] H.-T. Wu, "Adaptive analysis of complex data sets," Ph.D. dissertation, Princeton, 2011.

[17] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, "Time-frequency reassignment and synchrosqueezing: An overview," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 32–41, 2013.

[18] T. Oberlin, S. Meignen, and V. Perrier, "The Fourier-based synchrosqueezing transform," in *2014 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 315–319.

[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] D.-H. Pham and S. Meignen, "Second-order synchrosqueezing transform: The wavelet case, comparisons and applications," 2018. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01586372,2018

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

[22] 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.

[23] O. Aldimashki and A. Serbes, "Performance of chirp parameter estimation in the fractional fourier domains and an algorithm for fast chirp-rate estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 5, pp. 3685–3700, 2020.

[24] Y. Pantazis, O. Rosec, and Y. Stylianou, "Chirp rate estimation of speech based on a time-varying quasi-harmonic model," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 3985–3988.

[25] R. J. Baken and R. F. Orlikoff, *Clinical measurement of speech and voice*. Cengage Learning, 2000.

[26] J. M. Miramont, M. A. Colominas, and G. Schlotthauer, "Voice jitter estimation using high-order synchrosqueezing operators," *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, vol. 29, pp. 527–536, 2020.

[27] 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.

[28] 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.

[29] H. Holden, B. Øksendal, J. Ubøe, and T. Zhang, "Stochastic partial differential equations," in *Stochastic partial differential equations*. Springer, 1996, pp. 141–191.

[30] 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, March 2015.

[31] 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.

[32] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the lambertw function," *Advances in Computational mathematics*, vol. 5, no. 1, pp. 329–359, 1996.

[33] M. A. Colominas, S. Meignen, and D.-H. Pham, "Fully adaptive ridge detection based on STFT phase information," *IEEE Signal Processing Letters*, 2020.

[34] X. Zhu, H. Yang, Z. Zhang, J. Gao, and N. Liu, "Frequency-chirprate reassignment," *Digital Signal Processing*, vol. 104, p. 102783, 2020.

[35] B. Woldert-Jokisz. (2007) Saarbruecken voice database. [Online]. Available: http://stimmdb.coli.uni-saarland.de

[36] 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.

[37] 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.