High-order chirplet transform for efficient reconstruction of multicomponent signals with amplitude modulated crossing ridges

Timothée Maison, Fabrice Silva, Nathalie Henrich Bernardoni, Philippe Guillemain

High-order Chirplet Transform for Efficient Reconstruction of Multicomponent Signals with Amplitude Modulated Crossing Ridges

Timothée Maison[a,*], Fabrice Silva[a], Nathalie Henrich Bernardoni[b], Philippe Guillemain[a]

[a] *Aix-Marseille Univ., CNRS, Centrale Marseille, LMA, Marseille, France*
[b] *Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France*

Abstract

Multicomponent signals with crossing ridges, such as those encountered when measuring vocal tract resonances during singing, are challenging to analyze in time-frequency domain. The chirplet transform introduces the chirprate as a third dimension, extending the time-frequency domain to enable the separation of ridges. While existing methods assume weak amplitude modulations of signal components for the reconstruction process, a high-order chirplet transform is developed to accurately and efficiently retrieve amplitude modulation of crossing components, assuming that the instantaneous frequency of the components are already known. Analytical solving and numerical stability are obtained with a family of chirplet windows based on Hermite polynomials. The numerical results on simulated and real signals show the relevance and efficiency of the proposed method.

Submitted version, published article available at https://doi.org/10.1016/j.sigpro.2025.109887.

Keywords: Time-frequency, multicomponent signal, amplitude modulation, crossing ridges, chirplet transform

1. Introduction

The vocal tract, acting as an acoustical resonator, is one of the main functional levels of voice and speech production. Notably, the vocal tract resonances are greatly responsible for the shaping of the voice-source signal spectrum and the corresponding formants (implied, e.g., in the intelligibility of vowels). Estimation of resonant characteristics remains a challenge [1]. Formant analysis is still widely used, although it has been shown to rely on numerous biases [2]. Replicas can be used for direct comparison between measurements and modelling [3]. The need for *in-operando* characterization, i.e. on humans while they sing or speak, has led to the development of several methods, particularly the measurement of vocal tract response to an acoustic broadband excitation signal at the lips [4]. An approach is based on the generation of a perfectly known frequency-modulated chirp (linear or exponential) at the lips and the recording of the acoustical response of the resonating vocal tract [5]. Filtering of the resonances of the vocal tract modulates the amplitude of the chirp, with close local pairs of maxima and minima, but its instantaneous frequency is unchanged and known. Slow swept-sine excitation can be used for modal analysis, retrieving resonance characteristics [6], but *in-operando* measurement of vocal tract resonances requires fast excitation (ideally about a hundred milliseconds, currently around 1 s) with high chirp rate (to sweep the frequency range of the voice), increasing the amplitude modulation of the recorded chirp. In addition, during speaking or singing, the chirp is recorded along with the voice signal at the lips. The superposition of such non-stationary amplitude and frequency modulated components (chirp and voice harmonics) results in a multicomponent signal that has to be analyzed to estimate the vocal tract resonances.

Multicomponent signals can be studied by data driven analysis, as optimization problem to match a parametric model [7, 8]. Yet such an approach is not straight-forward as it depends on the optimization process and is usually time-consuming. Time-frequency (TF) methods provide powerful tools for analyzing multicomponent signals. Linear techniques such as the short-time Fourier transform or the continuous wavelet transform extend the one-dimensional time-series signal to the two-dimensional TF plane: the time and frequency characteristics of the signals can be observed simultaneously. However, a limitation lies in the Gabor-Heisenberg uncertainty principle, which reduces the readability of the TF representation [9]. Techniques are developed to recover TF signature of signal components. For instance, the reassignment method [10] offers a non-invertible post-processing improvement in energy concentration for the spectrogram. The synchrosqueezing transform [11] uses frequency reallocation and enables mode reconstruction for slightly modulated components. The synchroextracting transform [12] retains only the information related to the instantaneous frequency of the components. Detection of the ridges and instantaneous frequency (IF) of the components are required to perform the demodulation. Numerous papers propose IF estimators and extractors, e.g., [10, 13, 14], as it is a challenging

*Corresponding author
Email addresses: timothee.maison@ens-lyon.org (Timothée Maison), silva@lma.cnrs-mrs.fr (Fabrice Silva), nathalie.henrich@gipsa-lab.grenoble-inp.fr (Nathalie Henrich Bernardoni), guillemain@lma.cnrs-mrs.fr (Philippe Guillemain)

task. In this article, we focus on the reconstruction process of crossing amplitude-modulated components, assuming that the IFs are already known.

Extensions have been proposed to overcome the assumption of weak amplitude and frequency modulations usually made to apply those transforms. They take into account phase and amplitude modulations of second order [15] or higher order in the synchrosqueezing case [16], or in the synchroextracting case [17, 18], in order to propose robust IFs estimators and components reconstruction for AM-FM signals. These methods are incompatible with intersecting ridges and fail to separate components that are very close or that are even crossing in the TF plane.

Parametric TF analysis can overcome this issue, as is already the case in engineering applications [19]. The chirplet transform (CT) proposed by [20] allows the crossing components of a signal to be separated in the three-dimensional time-frequency-chirprate space using the chirprate parameter. Non-exhaustively, it has been applied for IFs estimation and retrieval of nonlinear signals [21], in reassignment [22], squeezing [23], synchroextracting [24], and mode retrieval [25]. Promising results have been obtained with a slow amplitude modulation constraint [26], or an equivalent approximation considering a locally constant amplitude linear chirp [27, 28, 25, 29]. Higher-order of amplitude and phase of the components have been taken into account for improving IFs estimates [30]. Yet the reconstruction process is still based on the approximation of a locally constant amplitude. Those analyzes, performing both IF estimates and components retrieval, often present the overall remaining error without specifying separately the error from IFs estimate and the one from reconstruction process. Particularly, the consequence of assuming a local weakly modulated amplitude or a constant amplitude is not addressed. It is therefore necessary to adopt a new approach, as such an approximation is not applicable in the context of this paper (i.e., in the context of the estimation of vocal tract resonances).

This paper aims at introducing a method relying on high-order chirplet transforms (HOCT) that separates and retrieves the different components of a signal, assuming that the components IFs are already known. In particular, the process of denoising is considered, that is, the estimation of the amplitude and the reconstruction of a given component in the presence of other ridges and/or noise in the signal. Focus is placed on the performance of the amplitude retrieval method, assuming that ridges are already detected.

This paper is structured as follows, proposing some theory on the signal processing method and numerical simulations. Some reminders on the short-time Fourier transform and its extension to the chirplet transform are given in Sec. 2. In Sec. 3, the amplitude estimation of a single component using the chirplet transform is explored, and a higher-order method based on Hermite polynomials to handle modulated amplitudes is introduced. The problem of separating crossing ridges in the TF plane is tackled in Sec. 4, firstly assuming time asymptoticity, and secondly in the general case of modulated amplitude components. Finally, Sec. 5 presents numerical results on simulated signals to show the relevance and the robustness of the proposed method against Gaussian noise. The method is applied on real signals in Sec. 6. For the sake of brevity, only the first step of the application of the HOCT method to vocal tract resonance measurement is presented in this document, even though it has already been successfully carried out [31].

2. Chirplet transform

2.1. The short-time Fourier transform

The *short-time Fourier transform* (STFT) can be defined for a function $s \in L^1(\mathbb{R})$ and a window w in the Schwartz class as a function of delay τ and angular frequency ω (see, e.g., [9]):

$$\begin{aligned} V^w(\tau,\omega) &= \int_{\mathbb{R}} s(t')w(t'-\tau)^* e^{-j\omega(t'-\tau)}dt' \\ &= \int_{\mathbb{R}} s(\tau+t)\left[w(t)e^{j\omega t}\right]^* dt \end{aligned} \tag{1}$$

where $*$ denotes the complex conjugate. The value $V^w(\tau,\omega)$ can be interpreted as the inner product of s shifted by τ and the TF atom defined as the restriction of a sine wave (with constant angular frequency ω) by the window w.

2.2. Extension for chirp atoms

As suggested in [32], atoms with linearly varying instantaneous (angular) frequency $\omega + \beta t$ (i.e., quadratic phase $\omega t + \beta t^2/2$) are now considered, so that the *chirplet transform* (CT) is defined as

$$\begin{aligned} L^w(\tau,\omega,\beta) &= \int_{\mathbb{R}} s(\tau+t)\left[w(t)e^{j(\omega t+\frac{1}{2}\beta t^2)}\right]^* dt \\ &= \int_{\mathbb{R}} s(\tau+t)\left[w_\beta(t)e^{j\omega t}\right]^* dt = V^{w_\beta}(\tau,\omega) \end{aligned} \tag{2}$$

where the so-called *chirplet* $w_\beta(t) = w(t)e^{j\frac{1}{2}\beta t^2}$ accounts for both the envelop w and the chirp rate β (see, e.g., [20]).

The Gaussian window $g^\sigma(t) = e^{-t^2/2\sigma^2}$ is used throughout this paper. The related chirplet is also a Gaussian function:

$$g^\sigma_\beta(t) = e^{-\frac{1}{2}t^2\left(1/\sigma^2 - j\beta\right)}.$$

The Fourier transform of the chirplet g^σ_β is thus

$$G^\sigma_\beta(\omega) = \mathcal{F}(g^\sigma_\beta)(\omega) = \frac{\sigma\sqrt{2\pi}}{\sqrt{1-j\beta\sigma^2}} \exp\left(-\frac{\omega^2\sigma^2}{2\left(1-j\beta\sigma^2\right)}\right).$$

A multicomponent signal (MCS) is considered as $s(t) = \sum_{p=1}^P s_p(t)$ where each component can be written as $s_p(t) = A_p(t)\exp\left(j\varphi_p(t)\right)$ with $A_p(t)$ the instantaneous amplitudes (assumed to be continuously differentiable), $\varphi_p(t)$ the phases (assumed to be at least twice continuously differentiable) and the instantaneous angular frequencies $\varphi'_p(t)$. The chirplet transform of the MCS is thus the sum $L^{g^\sigma}(\tau,\omega,\beta) = \sum_{p=1}^P L_p^{g^\sigma}(\tau,\omega,\beta)$ with

$$L_p^{g^\sigma}(\tau,\omega,\beta) = \int_{\mathbb{R}} A_p(\tau+t)\left[g^\sigma_\beta(t)e^{j\left(\omega t-\varphi_p(\tau+t)\right)}\right]^* dt \tag{3}$$

It is important to note that strict amplitude/phase separation of the components is not required. In fact, the phase $\varphi_p(t)$ can be chosen to only retain the main trend, e.g. a linear (or quadratic) evolution corresponding to a constant carrier frequency (or a linear chirp, respectively) and the complex amplitude $A_p(t)$ accounting for the remaining argument fluctuations. This can be useful when dealing with resonances of dynamical systems where Kramers-Kronig relations (related to causality) imply that argument variations come along with the modulation of the amplitude.

Throughout the subsequent sections, the superscript σ is omitted in the L^{g^σ} notation, which becomes L^g.

3. Amplitude reconstruction for a single component signal with constant chirp rate

3.1. Time asymptotic signal

First a single component ($P = 1$, omitted subscript p in all Sec. 3) is considered and time asymptoticity is assumed, so that the instantaneous amplitudes $A(\tau + t)$ and the IF derivative $\varphi''(\tau + t)$ are constant on the effective support of the window $g(t)$ (σ for the Gaussian window), i.e., $\varphi(\tau+t) = \varphi(\tau)+\varphi'(\tau)t+\frac{1}{2}t^2\varphi''(\tau)$ and

$$\begin{aligned} L^g(\tau,\omega,\beta) &= A(\tau)e^{j\varphi(\tau)}\int_{\mathbb{R}} g(t)e^{-j\left((\omega-\varphi'(\tau))t+\frac{1}{2}(\beta-\varphi''(\tau))t^2\right)}dt \\ &= A(\tau)e^{j\varphi(\tau)}G_{\beta-\varphi''(\tau)}\left(\omega-\varphi'(\tau)\right). \end{aligned} \tag{4}$$

As a first consequence, the instantaneous amplitude $A_p(\tau)$ can be estimated with a single evaluation of the CT:

$$\begin{aligned} L^g(\tau,\varphi'(\tau),\varphi''(\tau)) &= A(\tau)e^{j\varphi(\tau)}G_0\left(0\right) \\ \Rightarrow \widetilde{A}(\tau) &= \frac{e^{-j\varphi(\tau)}}{\sigma\sqrt{2\pi}}L^g(\tau,\varphi'(\tau),\varphi''(\tau)). \end{aligned} \tag{5}$$

Such an evaluation requires knowing the ridge position, i.e., the instantaneous frequency φ' and its derivative φ'' at (sampled) times τ. These values can be provided by *a priori* knowledge or computed using instantaneous frequency estimates provided in the literature, for example, in [13].

The second consequence is shown in Fig. 1. For a given delay τ, the CT $L^g(\tau,\omega,\beta)$ can be represented in subspace (ω,β) leading to the so-called *bowtie* pattern [20]. The CT deviates from its maximum value $L^g(\tau,\varphi'(\tau),\varphi''(\tau))$ according to the factor $G_{\beta-\varphi''(\tau)}\left(\omega-\varphi'(\tau)\right)$, which has interesting features. For fixed β, it is a Gaussian distribution centered on $\omega = \varphi'(\tau)$ with a variance that increases as $\sigma^{-1}\sqrt{1+\sigma^4(\beta-\varphi''(\tau))^2}$ with a minimum value σ^{-1} for $\beta = \varphi''(\tau)$. This implies that the frequency span of a component in the classical STFT (corresponding to case $\beta = 0$) may be much larger than that obtained when the chirp rate of the chirped window and of the component match (for $\beta = \varphi'(\tau)$). This span also critically depends on the variance σ of the Gaussian window. This behavior is shown to be related to the fractional Fourier transform [33].

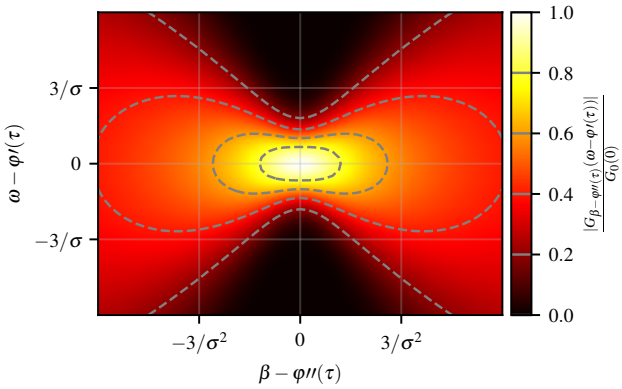

Figure 1: Evolution of $|G_{\beta-\varphi''(\tau)}(\omega-\varphi'(\tau))|$ as a function of ω and β.

3.2. Accounting for amplitude modulation with higher orders

The hypothesis of constant amplitude modulation (time asymptotic hypothesis) is now relaxed. The instantaneous amplitude can be represented as a truncated Taylor series, taking into account its modulation to an order $2K$

$$A(\tau+t) \underset{t\simeq 0}{=} \sum_{k=0}^{2K}\frac{t^k}{k!}A^{(k)}(\tau) + o(t^{2K+1}). \tag{6}$$

Such an expansion is often considered for high-order IF estimation [16, 17, 30]. Several terms now contribute to the CT $L^g(\tau,\omega,\beta)$ and disable the direct estimate of $A(\tau)$ without a zero-order truncation of the amplitude. In order to restore this possibility, a set of windows $(g_n)_{n\in[0,N]}$ is introduced. The corresponding CTs $L^{g_n}(\tau,\varphi'(\tau),\varphi''(\tau))$ are formulated as follows:

$$\begin{bmatrix} L^{g_0}(\tau,\varphi'(\tau),\varphi''(\tau)) \\ L^{g_1}(\tau,\varphi'(\tau),\varphi''(\tau)) \\ \vdots \\ L^{g_N}(\tau,\varphi'(\tau),\varphi''(\tau)) \end{bmatrix} = e^{j\varphi(\tau)}\mathbb{B}^{(g)}\begin{bmatrix} A(\tau) \\ A''(\tau) \\ \vdots \\ A^{(2K)}(\tau) \end{bmatrix} \tag{7}$$

with the matrix $\mathbb{B}^{(g)}$ (square if $N = K$) defined by

$$\mathbb{B}^{(g)}_{n,k} = \int_{\mathbb{R}}\frac{t^{2k}g_n(t)}{(2k)!}dt. \tag{8}$$

involving only even order derivatives if we consider only even windows g_n (i.e., $g_n(-t) = g_n(t)$). When choosing the latter as Gaussian multiplied by canonical monomials, i.e., $g_n(t) = (t/\sigma)^{2n}e^{-t^2/(2\sigma^2)}$, it follows that

$$\mathbb{B}^{(g)}_{n,k} = \sigma^{2k+1}\sqrt{2\pi}\frac{(2k+2n-1)!!}{(2k)!}$$

where !! denotes the double factorial.

The condition number of the matrix $\mathbb{B}^{(g)}_{n,k}$ increases dramatically with K, resulting in numerical issues for its inversion (or, equivalently, for the resolution of Eq. (7)). In order to avoid the full resolution of Eq. (7) for each delay τ, the set of windows that would produce the identity matrix $\mathbb{B} = \mathbb{I}_{K+1}$ could be considered. This would lead to a set of windows valid for a single truncation order K given. Changing the truncation order

would change the full set of windows. For instance, the first window g_0 would not even converge to a regular function when increasing K.

An incremental approach is preferred here, building a unique set of functions (denoted (h_n)) for which the truncation order K only changes the number of functions to consider. The set of functions

$$h_n(t) = \frac{1}{\sigma\sqrt{2\pi}} H_{2n}\left(\frac{t}{\sigma}\right) e^{-\frac{t^2}{2\sigma^2}} \tag{9}$$

is based on the even probabilist's Hermite polynomials defined as

$$H_{2n}(t) = e^{\frac{t^2}{2}} \frac{d^{2n}}{dx^{2n}} e^{-\frac{t^2}{2}}. \tag{10}$$

Such a definition ensures that $\mathbb{B}^{(h)}$ is an upper triangular Toeplitz matrix. For $N = K$, it is a square matrix whose inverse is also an upper triangular Toeplitz matrix and can be analytically computed:

$$\mathbb{B}^{(h)}_{n,k} = \frac{\sigma^{2k}}{2^{k-n}(k-n)!} \text{ if } k \geq n, 0 \text{ otherwise,}$$
$$\left(\mathbb{B}^{(h)}\right)^{-1}_{k,n} = \frac{(-1)^{n-k}}{\sigma^{2k}2^{n-k}(n-k)!} \text{ if } n \geq k, 0 \text{ otherwise,}$$

so the amplitude $A(\tau)$ can simply be estimated from the first line of $(\mathbb{B}^{(h)})^{-1}$ (i.e., $k = 0$):

$$\widetilde{A}(\tau) = e^{-j\varphi(\tau)} \sum_{n=0}^{N} \frac{(-1)^n}{2^n\, n!} L^{h_n}(\tau, \varphi'(\tau), \varphi''(\tau)).$$

This series clearly identifies the respective contribution of each CT when increasing the truncation order K.

However, $\mathbb{B}^{(g)}_{n,k}$ as an upper triangular matrix is not optimum from the computational cost point of view. Thus, we abstract the Taylor series expansion (Eq. (6)) and consider a more generic representation of the amplitude modulation

$$A(\tau + t) = \sum_{k=0}^{K} \alpha_k(\tau) \Phi_k\left(\frac{t}{\sigma}\right) \tag{11}$$

based on the family of functions Φ_k of the reduced time and where α_k are the coefficient to be determined ($\Phi_k(\theta) = \sigma^k\theta^k/k!$ and $\alpha_k(\tau) = A^{(k)}(\tau)$ for the Taylor series in Eq. (6)). For a set of functions f_n to be defined hereafter, Eq. (7) now writes

$$\begin{bmatrix} L^{f_0}(\tau, \varphi'(\tau), \varphi''(\tau)) \\ L^{f_1}(\tau, \varphi'(\tau), \varphi''(\tau)) \\ \vdots \\ L^{f_N}(\tau, \varphi'(\tau), \varphi''(\tau)) \end{bmatrix} = e^{j\varphi(\tau)} \mathbb{B}^{(f)} \begin{bmatrix} \alpha_0(\tau) \\ \alpha_1(\tau) \\ \vdots \\ \alpha_K(\tau) \end{bmatrix} \tag{12}$$

with the matrix $\mathbb{B}^{(f)}$ defined by

$$\mathbb{B}^{(f)}_{n,k} = \int_{\mathbb{R}} \Phi_k\left(\frac{t}{\sigma}\right) f_n(t) dt. \tag{13}$$

Taking advantage of the orthogonality property of the Hermite polynomials, we consider those for both the set of windows f_n (see Fig. 2), with a normalization slightly different from h_n,

$$f_n(t) = \frac{1}{\sigma\sqrt{2\pi}(2n)!} H_{2n}\left(\frac{t}{\sigma}\right) e^{-\frac{t^2}{2\sigma^2}}. \tag{14}$$

and for the expansion of the amplitude modulation Φ_k (with a shift such that $A(\tau)$ reduces to $\alpha_0(\tau)$)

$$\Phi_k(\theta) = \begin{cases} H_0(\theta) = 1 & \text{if } k = 0, \\ H_{2k}(\theta) - H_{2k}(0) & \text{otherwise.} \end{cases} \tag{15}$$

The first-order chirplet atom based on the Hermite polynomials is the Gaussian window. Higher orders of polynomials correspond in the TF plane to two complementary Gaussian chirplets separated by a fixed frequency, resulting in beats in the time domain. The higher the order, the higher the beat frequency in the time domain and the more distant the complementary chirplets are in frequency in the TF plane.

The matrix $\mathbb{B}^{(f)}$ is almost empty except on the first line and on the main diagonal, which can be calculated as

$$\mathbb{B}^{(f)}_{0,k>0} = (-1)^{k+1}(2k-1)!! \text{ and } \mathbb{B}^{(f)}_{n,n} = 1 \tag{16}$$

Its inverse can also be analytically computed in the square case ($N = K$), avoiding potential numerical issues, so that, even without any time asymptoticity hypothesis on the instantaneous amplitude, the latter being estimated by

$$\widetilde{A}(\tau) = e^{-j\varphi(\tau)}[L^{f_0}(\tau, \varphi'(\tau), \varphi''(\tau)) - \sum_{n=0}^{N}(-1)^n(2n-1)!!L^{f_n}(\tau, \varphi'(\tau), \varphi''(\tau))] \tag{17}$$

extending Eq. (5). The numerical results on reconstruction and convergence are presented in Sec. 5.

3.3. Complexity and efficiency

Regarding computational cost, the choice of canonical windows (Taylor expansion) Eqs. (6), (7), (8) requires solving a linear system with a full matrix (no particular property), that is, $O(N^3)$ multiplications with a direct algorithm (using Gaussian elimination). The set of Hermite functions (Eq. (9)) leads to an upper triangular matrix with explicit inverse so that solving the linear system corresponds to $O(N^2)$ multiplications (or only $O(N)$ to retrieve the sole $A(\tau)$). The last configuration Eqs. (15), (14), (12), (16) reduces the number of multiplications to $O(N)$, which is equivalent to the estimation of $A(\tau)$ but clearly better than the previous choices for the full AM expansion. In addition, the costs involved in a frame-based analysis can be compared. First of all, the computation of the windows (Gaussian, or Hermitian Eqs.(9), (14)) is obviously factorized across frames and only performed once, and is thus a negligible overhead. For each frame, the cost of the $N + 1$ evaluations of the CT along the ridge is independent of the choice of windows (the number of multiplications corresponds to the length of the frame), so that the cost of inversion involved in the estimate $\widetilde{A}(\tau)$ is really a key point.

4. Extension to multicomponent signal with crossing ridges

4.1. Time asymptotic components

A signal with two independent components $s(t) = s_1(t) + s_2(t)$ ($P = 2$) is now considered. We assume that $s_1(t) =$

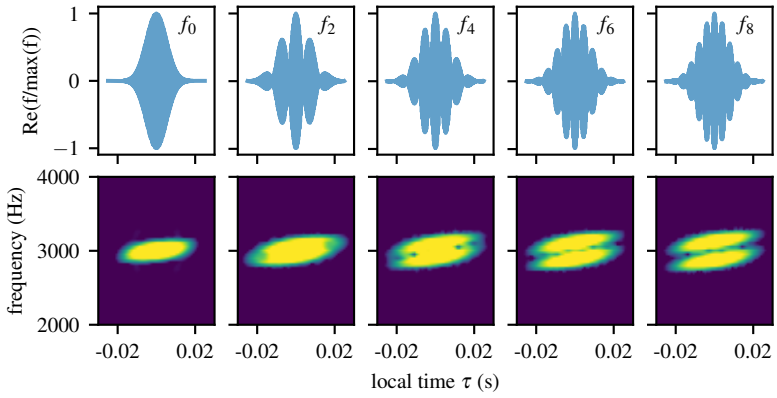

Figure 2: Time representation and magnitude spectrograms of normalized chirplet atoms based on Hermite polynomials ($f_n/max(f_n)$) defined Eq. (14), used for taking into account amplitude modulation, centered on frequency $f = 3\,\text{kHz}$ for chirprate $\beta = 6\,\text{kHz}\,\text{s}^{-1}$.

$A_1(t)e^{j\varphi_1(t)}$ and $s_2(t) = A_2(t)e^{j\varphi_2(t)}$ are time asymptotic (constant instantaneous amplitude and frequency variation on the effective support of the window $g(t)$). Relying on Eq. (4), the chirplet transform can then be written as

$$\begin{aligned} L^g(\tau,\omega,\beta) = A_1(\tau)e^{j\varphi_1(\tau)}G_{\beta-\varphi_1''(\tau)}\left(\omega-\varphi_1'(\tau)\right) \\ + A_2(\tau)e^{j\varphi_2(\tau)}G_{\beta-\varphi_2''(\tau)}\left(\omega-\varphi_2'(\tau)\right) \end{aligned} \tag{18}$$

In line with Sec. 3, we focus on the instantaneous amplitudes $A_p(\tau)$ under the assumption that the location of both ridges is known (i.e. instantaneous phase, frequency, and its derivative). The amplitudes can then be estimated using two evaluations of the CT transform at locations (τ,ω_1,β_1) and (τ,ω_2,β_2):

$$\begin{bmatrix} A_1(\tau)e^{j\varphi_1(\tau)} \\ A_2(\tau)e^{j\varphi_2(\tau)} \end{bmatrix} = \mathbb{S}^{(g)\,-1}\begin{bmatrix} L^g(\tau,\omega_1,\beta_1) \\ L^g(\tau,\omega_2,\beta_2) \end{bmatrix} \tag{19}$$

with

$$\mathbb{S}^{(g)}_{ij} = G_{\beta_i-\varphi_j''(\tau)}\left(\omega_i-\varphi_j'(\tau)\right) \tag{20}$$

Without any other knowledge, an obvious choice is to adapt the chirplets to the ridges:

$$(\omega_1,\beta_1) = (\varphi_1'(\tau),\varphi_1''(\tau)) \text{ and } (\omega_2,\beta_2) = (\varphi_2'(\tau),\varphi_2''(\tau))$$

which leads to a self-adjoint (Hermitian) matrix with maximal contributions on the diagonal

$$\begin{cases} \mathbb{S}^{(g)}_{11} = \mathbb{S}^{(g)}_{22} = G_0(0) = \sigma\sqrt{2\pi}, \\ \mathbb{S}^{(g)}_{12} = (\mathbb{S}^{(g)}_{21})^* = G_{\Delta\varphi''}\left(\Delta\varphi'\right) \end{cases}$$

denoting $\Delta\varphi' = \varphi_1'(\tau) - \varphi_2'(\tau)$ and $\Delta\varphi'' = \varphi_1''(\tau) - \varphi_2''(\tau)$ the distances between the two ridges. However, the matrix becomes ill-conditioned when two ridges cross with a small angle ($|\Delta\varphi''| \ll \sigma^{-2}$).

Another inconvenience is that the estimates are not robust to uncertainty on the location of one of the ridges. In fact, error analysis shows that the estimates are maximally sensitive to errors on $\varphi_i'(\tau)$ when the frequency distance between the ridges is about σ^{-1}. This is graphically depicted in Fig. 1 by the greatest slope of $|G(\omega,\beta)|$ just above and below the center of the *bowtie*: a small frequency error can then lead to a strong bias in the matrix $\mathbb{S}^{(g)}$ and in the estimates of the amplitudes.

An alternative choice can be made when one of the ridges is perfectly localized, permitting some uncertainty on the other one. Setting $\omega_1 = \omega_2 = \varphi_1'(\tau)$, $\beta_1 = \varphi_1''(\tau)$ and β_2 such as $|\beta_2 - \varphi_i''(\tau)| > \sigma^{-2}$ ensures that the matrix $\mathbb{S}^{(g)}$ is well-conditioned. For monotonic chirps (i.e., non vanishing $\varphi_1''(\tau)$), one can choose $\beta_2 = -\varphi_1''(\tau)$, leading to

$$\mathbb{S}^{(g)} = \begin{bmatrix} G_0(0) & G_{\Delta\varphi''}\left(\Delta\varphi'\right) \\ G_{-2\varphi_1''(\tau)}(0) & G_{-2\varphi_1''(\tau)+\Delta\varphi''}\left(\Delta\varphi'\right) \end{bmatrix} \tag{21}$$

In the particular process of denoising a given component $s_1(t)$ (i.e., when the main objective is to recover only $A_1(\tau)$), the instantaneous phase of the other component $\varphi_2(\tau)$ does not need to be known. The numerical results on denoising performance and robustness are presented in Sec. 5.

4.2. Amplitude-modulated components

Finally, the case of a multicomponent signal with amplitude modulation is considered, illustrating the complete high-order chirplet transforms (HOCT) method. For the sake of conciseness, the signal has $P = 2$ components such that $s(t) = s_1(t) + s_2(t) = A_1(t)e^{j\varphi_1(t)} + A_2(t)e^{j\varphi_2(t)}$. The time asymptotic assumption on $s_1(t)$ and $s_2(t)$ is relaxed. As in Sec. 3.2, their amplitudes are expanded on Hermite polynomials:

$$\begin{aligned} A_1(\tau+t) &\simeq \sum_{k=0}^{K}\alpha_k^1(\tau)\Phi_k\left(\frac{t}{\sigma}\right) \\ A_2(\tau+t) &\simeq \sum_{k=0}^{L}\alpha_k^2(\tau)\Phi_k\left(\frac{t}{\sigma}\right) \end{aligned}$$

with the basis functions Φ_k defined in Eq. (15), and the CTs are computed with the set of windows $f_n(t)$ defined in Eq (14). Extending the matrices used in the AM reconstruction (Sec. 3.2) and in the ridge separation (Sec. 4.1), one can then write the

complete matrix system :

$$\begin{bmatrix} L^{f_0}(\tau,\omega_1,\beta_1) \\ \vdots \\ L^{f_N}(\tau,\omega_1,\beta_1) \\ \hline L^{f_0}(\tau,\omega_2,\beta_2) \\ \vdots \\ L^{f_M}(\tau,\omega_2,\beta_2) \end{bmatrix} = \left[\begin{array}{c|c} \mathbb{B}^{(f)} & \mathbb{C}^{(f)} \\ \hline \mathbb{D}^{(f)} & \mathbb{E}^{(f)} \end{array}\right] \begin{bmatrix} \alpha_0^1 e^{j\varphi_1(\tau)} \\ \vdots \\ \alpha_K^1 e^{j\varphi_1(\tau)} \\ \hline \alpha_0^2 e^{j\varphi_2(\tau)} \\ \vdots \\ \alpha_L^2 e^{j\varphi_2(\tau)} \end{bmatrix} \tag{22}$$

The various blocks derive from a common quantity $\mathbb{F}$:

$$\begin{aligned} \mathbb{B}^{(f)} &= \mathbb{F}\left(\omega_1 - \varphi_1'(\tau), \beta_1 - \varphi_1''(\tau)\right) \\ \mathbb{C}^{(f)} &= \mathbb{F}\left(\omega_1 - \varphi_2'(\tau), \beta_1 - \varphi_2''(\tau)\right) \\ \mathbb{D}^{(f)} &= \mathbb{F}\left(\omega_2 - \varphi_1'(\tau), \beta_2 - \varphi_1''(\tau)\right) \\ \mathbb{E}^{(f)} &= \mathbb{F}\left(\omega_2 - \varphi_2'(\tau), \beta_2 - \varphi_2''(\tau)\right) \end{aligned}$$

with generic term

$$\mathbb{F}_{n,l}(\Delta\omega, \Delta\beta) = \int_{\mathbb{R}} \Phi_l\left(\frac{t}{\sigma}\right) f_n(t) e^{-j\Delta\omega t} e^{-j\frac{1}{2}\Delta\beta t^2} dt. \tag{23}$$

The integral can be evaluated as

$$\mathbb{F}_{n,l}(\Delta\omega, \Delta\beta) = \begin{cases} \mathbb{G}_{2n,0}(\Delta\omega, \Delta\beta) \text{ if } l = 0 \\ \mathbb{G}_{2n,2l}(\Delta\omega, \Delta\beta) - H_{2l}(0)\mathbb{G}_{2n,0}(\Delta\omega, \Delta\beta) \end{cases} \tag{24}$$

with

$$\begin{aligned} \mathbb{G}_{2n,2l}(\Delta\omega, \Delta\beta) &= \int_{\mathbb{R}} H_{2l}\left(\frac{t}{\sigma}\right) \frac{1}{\sigma\sqrt{2\pi}(2n)!} H_{2n}\left(\frac{t}{\sigma}\right) e^{-j\Delta\omega t} e^{-j\frac{1}{2}\Delta\beta t^2} \\ &= \frac{\zeta_{\Delta\beta}}{(2n)!} e^{-\frac{1}{2}(\sigma\Delta\omega\zeta_{\Delta\beta})^2} \sum_{p=0}^{2\min(l,n)} \frac{(2n)!(2l)!}{p!(2n-p)!(2l-p)!} \zeta_{\Delta\beta}^{2p} \\ &\times \left(1 - \zeta_{\Delta\beta}^2\right)^{n+l-p} H_{2n-p}\left(\frac{j\sigma\Delta\omega\zeta_{\Delta\beta}^2}{\sqrt{1-\zeta_{\Delta\beta}^2}}\right) H_{2l-p}\left(\frac{j\sigma\Delta\omega\zeta_{\Delta\beta}^2}{\sqrt{1-\zeta_{\Delta\beta}^2}}\right) \end{aligned} \tag{25}$$

with $\zeta_{\Delta\beta} = 1/\sqrt{1 + j\sigma^2\Delta\beta}$. See Appendix B for key calculus elements.

The two amplitude-modulated components case Eq. (22) or even a general case with more components could be solved as a system of linear equations with algebra methods.

However, in line with Sec. 4.1 we will focus here on the particular process of denoising the first component. We assume a time-asymptotic second component ($M = L = 0$), which leads $\mathbb{C}_n = \mathbb{F}_{n,0}(\Delta\varphi'(\tau), \Delta\varphi''(\tau)) = \mathbb{G}_{2n,0}(\Delta\varphi'(\tau), \Delta\varphi''(\tau))$ to be a column vector, $\mathbb{D}_k = \mathbb{F}_{0,l}(0, -2\varphi''(\tau))$ a row vector, and $\mathbb{E}$ a scalar. Moreover, we set $\omega_1 = \omega_2 = \varphi_1'(\tau)$, $\beta_1 = \varphi_1''(\tau)$ and $\beta_2 = -\varphi_1''(\tau)$ so that $\mathbb{B} = \mathbb{F}_{n,l}(0,0)$ is the same matrix as Eq. (16), ensuring its easy inversion, and $\mathbb{E} = \mathbb{S}_{22} = G_{-2\varphi_1''(\tau)+\Delta\varphi''}(\Delta\varphi')$ (see Eq. (21)). The reverse system can then be analytically computed, using the Schur complement of $\mathbb{B}$ defined by the scalar $u = \mathbb{E} - \mathbb{D}\mathbb{B}^{-1}\mathbb{C}$, and its inverse is trivial as long as $u \neq 0$.

$$\begin{bmatrix} \alpha_0^1 e^{j\varphi_1(\tau)} \\ \vdots \\ \alpha_K^1 e^{j\varphi_1(\tau)} \\ \hline \alpha_0^2 e^{j\varphi_2(\tau)} \end{bmatrix} = \mathbb{X} \begin{bmatrix} L^{f_0}(\tau,\omega_1,\beta_1) \\ \vdots \\ L^{f_K}(\tau,\omega_1,\beta_1) \\ \hline L^{f_0}(\tau,\omega_2,\beta_2) \end{bmatrix} \tag{26}$$

with

$$\mathbb{X} = \begin{bmatrix} \mathbb{B}^{-1} + \mathbb{B}^{-1}\mathbb{C}u^{-1}\mathbb{D}\mathbb{B}^{-1} & -\mathbb{B}^{-1}\mathbb{C}u^{-1} \\ -u^{-1}\mathbb{D}\mathbb{B}^{-1} & u^{-1} \end{bmatrix}. \tag{27}$$

5. Numerical results

In this section, the relevance and performance of the proposed method are evaluated on synthetic signals, with a denoising point of view, i.e. focusing on the reconstruction of one given ridge of a MCS among others, in line with Eq. (26). The reconstruction error between the original component $s_p(t)$ and its estimate $\tilde{s}_p(t)$ is evaluated by the output Signal-to-Noise Ratio

$$\mathrm{SNR}_{\mathrm{out}}(s_p, \tilde{s}_p) = 20\log_{10}\left(\frac{\|s_p\|_2}{\|s_p - \tilde{s}_p\|_2}\right) \tag{28}$$

This ratio is a global error measure directly related to the relative root mean square error (RRMSE). Comparing signals (s_p vs $\tilde{s}_p$) or amplitudes (A_p vs $\tilde{A}_p$) is equivalent since *a priori* knowledge of their common phase is assumed. The global error measure may be complemented by a local relative error

$$\varepsilon(\tau) = \left|\frac{\tilde{A}_p(\tau)}{A_p(\tau)} - 1\right| \tag{29}$$

that helps to identify the conditions of worst-case error.

All signals last for 1 s. The sampling frequency is 44.1 kHz. Analyses are carried out using frames of length 50 ms with 98 % overlap. The Gaussian width is set to $\sigma = 5.2$ ms so that the Gaussian window decreases to 10^{-5} at the two ends of the frame.

5.1. Amplitude-modulated chirp

The HOCT method is first illustrated with a single-component amplitude-modulated signal $s_1(t) = A_1(t)e^{j\varphi_1(t)}$ (in line with Sec 3.2), displayed in Fig. 3, with:

$$\begin{aligned} A_1(t) &= 1 + 0.5\cos(2\pi\, 20t) \\ \varphi_1(t) &= 2\pi\left(100t + 6000t^2/2\right) \end{aligned} \tag{30}$$

The window length ($T_w = 50$ ms) is in the same range as the modulation period of the signal amplitude ($1/20 = 50$ ms) so the time asymptotic hypothesis is clearly not verified. The chirp rate $\varphi_1''(t) = 2\pi \times 6000 = 37.7 \times 10^3$ rad s^{-2} is comparable to σ^{-2} which induces the blurring of the ridge in the STFT spectrogram in Fig. 3b.

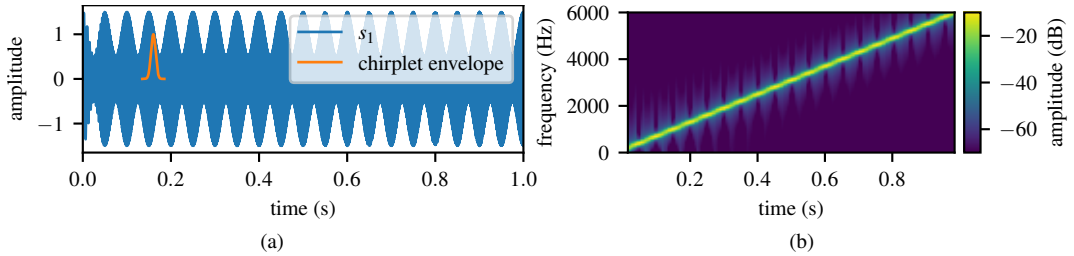

Figure 3: Waveform (a) of s_1 and chirplet envelope centered on $\tau = 0.16$ s, and spectrogram (b) of s_1.

5.1.1. *Relevance of the AM method*

The amplitude of the signal is estimated by the method described in Sec 3.2 considering the perfect knowledge of the phase $\varphi(t)$, for two numbers of computed higher-order moments $N = 0$ (Eq. (5), i.e. assuming a time asymptotic signal) and $N = 5$ (Eq. (17), i.e. taking into account up to the $2K = 10^{\text{th}}$ derivative of the amplitude), respectively. The original amplitude and the estimates are shown in Fig. 4.

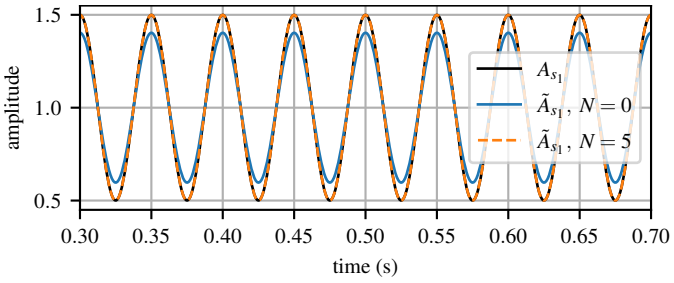

Figure 4: Original A_1 and estimated amplitude $\tilde{A}_1$ for $N = 0$ and $N = 5$ in Eq (17) in the noise-free case, zoom in $t = 0.3 - 0.7$ s.

The estimated amplitude $\widetilde{A_1}(t)$ without HOCT ($N = 0$) clearly fails in reproducing the variation of amplitude $A_1(t)$, with a corresponding $\text{SNR}_{\text{out}}(s_1, \widetilde{s_1}) = 23.8$ dB. In comparison, the HOCT analysis ($N = 5$) leads in this case to a perfect reconstruction of the signal amplitude modulation ($\text{SNR}_{\text{out}} = 91.9$ dB). This result means that a polynomial of degree $2K = 10$ (with $K = N$) is sufficient (and possibly overfitting) to correctly reconstruct the amplitude modulation over the given time range, here a full period of the amplitude modulation.

5.1.2. *Robustness to Gaussian noise*

To evaluate the robustness of the method to divergence from the MCS model, white Gaussian noise is added to $s_1(t)$ (Eq. (30)). Estimates are made as before. Input signal-to-noise ratio

$$\text{SNR}_{\text{in}}(s, \eta) = 20 \log_{10} (\|s\|_2 / \|\eta\|_2) \tag{31}$$

varies from $+\infty$ (noiseless case) to 0 (very noisy case). The results are shown in Table 1, with a mean statistic computed on 10 repetitions. For all SNR_{in} tested, the HOCT method performs better than the zero-order one. SNR_{out} decreases as the SNR_{in} decreases, which indicates the method's sensibility to noise. In the very noisy case ($\text{SNR}_{\text{in}} = 0$), the improvement made by the HOCT method tends to vanish.

Table 1: Performance of amplitude estimate of a single component with white Gaussian noise (SNR_{out} in dB) without ($N = 0$) and with HOCT ($N = 5$).

SNR_{in} (dB)	∞	20	10	3	0
$N = 0$	23.8	23.8	23.7	23.3	22.8
$N = 5$	91.9	43.8	33.8	26.9	23.9

5.1.3. *Convergence analysis*

To evaluate the convergence of the estimated amplitude with respect to the window length and to number N of computed moments, a modulation metric relative to the effective window length σ is defined as:

$$m_A = \max_t \left[\frac{\sigma^2 f_A}{A(t)} \frac{dA(t)}{dt} \right] \tag{32}$$

with f_A the modulation frequency (in previous paragraph: $f_A = 20$ Hz). The metric takes into account the number of amplitude oscillations seen by the chirplet window. For example, the amplitude modulation presented in Fig. 3 ($f_A = 20$ Hz) is related to $m_A \simeq 3.9 \times 10^{-5}$. Increasing σ, f_A, or amplitude modulation leads to a relatively similar analysis. Here, the frequency of the amplitude modulation of signal s_1 is varying from $f_A = 0$ Hz (no AM) to $f_A = 70$ Hz (very strong AM) with 10 Hz steps. The signal amplitude is then estimated for a number of higher-order CTs from $N = 0$ to $N = 15$ for each step, and SNR_{out} are computed to evaluate the quality of the estimates. The convergence results are shown in Fig. 5.

For a given amplitude modulation, increasing the number of higher-order moments computed leads to a better SNR_{out} that converges around 90 dB (perfect reconstruction). Note that for real signals which generally do not present very strong amplitude modulations, it should not be necessary to exceed the order $N = 10$ to obtain sufficient results in most of the cases ($\text{SNR}_{\text{out}} > 60$ dB, i.e. magnitude of reconstruction errors 1000 times smaller than signal amplitude).

5.2. *Separation of crossing ridges with constant amplitude*

To illustrate the classical separation result of the CT, the constant amplitude case with a two-components signal $s(t) =$

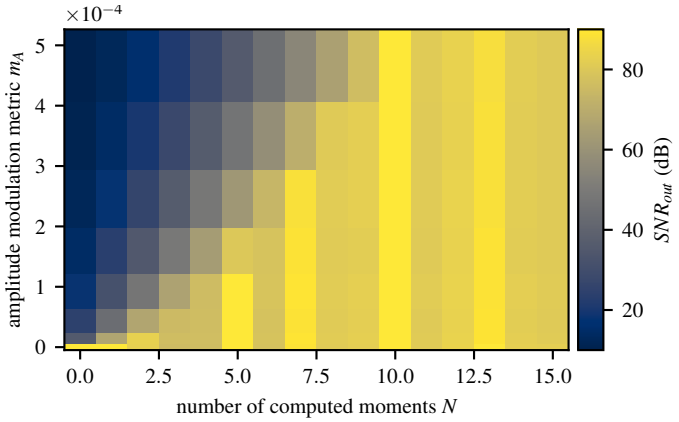

Figure 5: Convergence analysis: $\mathrm{SNR}_{\mathrm{out}}$ as a function of the number of computed higher-order CTs and of the amplitude modulation metric m_A of the signal s_1.

$s_2(t) + s_3(t) = A_2(t)e^{j\varphi_2(t)} + A_3(t)e^{j\varphi_3(t)}$ is considered, with:

$$\begin{aligned} A_2(t) &= A_3(t) = 1 \\ \varphi_2(t) &= 2\pi\left(100t + 6000t^2/2\right) = \varphi_1(t) \\ \varphi_3(t) &= 2\pi\left(2100t + 2000t^2/2\right) \end{aligned} \tag{33}$$

such as the ridges cross at $\omega = 2600$ Hz, at $t = 0.5$ s (see Fig. 6).

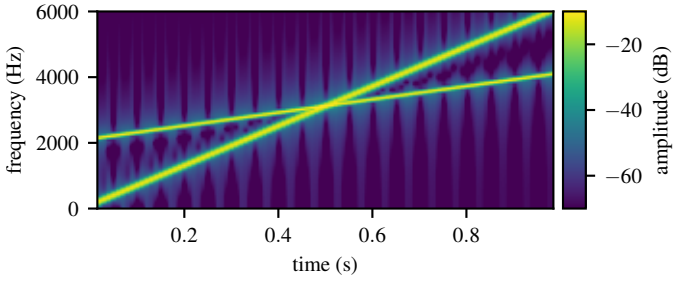

Figure 6: Spectrogram of signal $s(t) = s_2(t) + s_3(t)$.

The separation method described in Sec. 4.1 is applied in the denoising case (same as Eq. (21)) with the CT parameters $\omega_1 = \omega_2 = \varphi_2'(\tau)$ and $\beta_1 = -\beta_2 = \varphi_2''(\tau)$. The relative error on the estimated amplitude $\widetilde{A}_{s_2}$ with and without the separation method is presented in Fig. 7. From a practical point of view, the separation method using the CT with parameters (ω_2, β_2) is only applied when the contribution factor of s_3 to the CTs reaches a significant threshold, i.e., for $e^{-\sigma^2\left(\varphi_3'(\tau)-\varphi_2'(\tau)\right)^2/2} > 10^{-20}$.

The relative error on the estimated amplitude A_2 reaches a maximum at the crossing point between the two components in the case of an analysis with only one chirplet transform (Eq. (5)), indicating that it fails in the separation process. The use of the separation method (Eq. (21)) leads to a perfect retrieval of the amplitude ($\forall t, \varepsilon(t) \simeq 10^{-6}$). The reconstruction signal $\widetilde{s_1}$ can be considered perfect with the separation method regardless of the amplitudes of the ridges: for example, with $\mathrm{SNR}_{\mathrm{in}}(s_2, s_3) = -20$ dB, $\mathrm{SNR}_{\mathrm{out}}(s_2, \tilde{s}_2) = 109.8$ dB (relative error inferior to 10^{-5}).

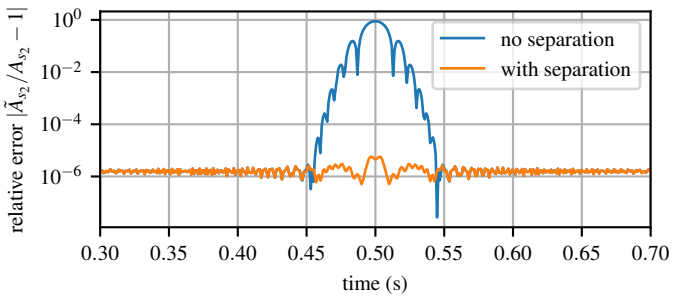

Figure 7: Relative error on estimated amplitude $|\tilde{A}_{s_2}/A_{s_2} - 1|$ with and without (w/o) separation method applied on $s_2 + s_3$, zoom in $t = 0.3 - 0.7$ s.

5.2.1. *Robustness to Gaussian noise*

White Gaussian noise is added to s with $\mathrm{SNR}_{\mathrm{in}}(s, \eta)$ varying from $+\infty$ (noise-free) to 0 (very noisy). Table 2 shows the mean values obtained on 10 realizations. For all tested $\mathrm{SNR}_{\mathrm{in}}$, taking both components into account in the separation method performs better than using a single chirplet transform.

Table 2: Performance of amplitude estimate ($\mathrm{SNR}_{\mathrm{out}}$) of crossing ridges with white Gaussian noise with or without (w/o) separation.

$\mathrm{SNR}_{\mathrm{in}}$ (dB)	∞	20	10	3	0
w/o separation (Eq. (5))	19.7	19.6	19.6	19.5	19.2
with separation (Eq. (21))	115.4	48.9	39.0	32.0	29.0

5.3. *Crossing ridges separation with amplitude modulation*

Amplitude modulation is now reintroduced considering the signal $s(t) = s_1(t) + s_3(t)$, with s_1 and s_3 previously described in Eqs. (30) and (33). The amplitude of s_1 is estimated using Eq. (26), i.e., accounting both for ridge separation and amplitude modulation, as the results are compared to those obtained in the other configurations: with or without ridge separation, zero-order ($N = 0$) or higher order ($N = 5$) computed CTs. The relative error made on the estimated amplitude for each configuration is shown in Fig. 8 and the associated $\mathrm{SNR}_{\mathrm{out}}(s_2, \tilde{s}_2)$ are listed in Table 3.

Table 3: Performances of the amplitude estimate ($\mathrm{SNR}_{\mathrm{out}}$) in the noise-free case.

	$N = 0$	$N = 5$
w/o separation	19.4 dB [Eq. (5)]	16.3 dB [Eq. (17)]
with separation	23.4 dB [Eq. (21)]	64.6 dB [Eq. (26)]

Outside of the frames in the vicinity of the ridges crossing ($\tau \simeq 0.5$ s), the relative error on estimated amplitude A_1 is smaller (between 10^{-5} and 10^{-4}) when modulation amplitude is taken into account ($N = 5$). When ridges are crossing, accounting for AM but not for nearby ridge (green solid curve, Eq. (17)) may perform worse than the method only intended for ridge separation (orange dashed curve, Eq. (21)). This is also visible in Table 3. Accounting for amplitude modulation with

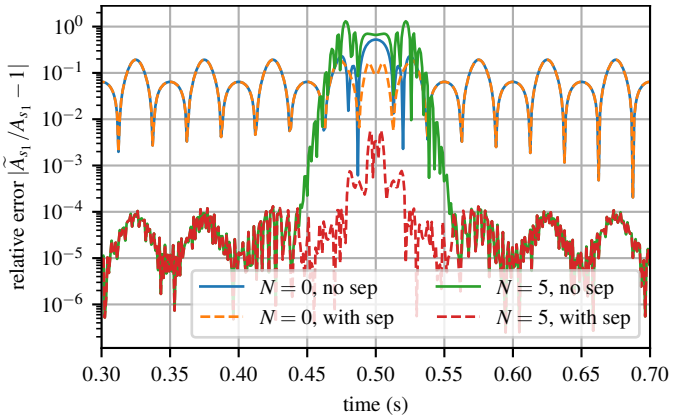

Figure 8: Relative error on estimated amplitude $|\widetilde{A}_{s_1}/A_{s_1} - 1|$ without (w/o sep) and with (with sep) separation method applied on $s_1 + s_3$, and with number of higher-order CTs computed to account for amplitude modulation $N = 0$ and $N = 5$, zoom in $t = 0.3 - 0.7$ s.

the separation method (Eq. (26)) provides the best results even if the separation is still not perfect (2×10^{-3} maximal relative error attained near ridge crossing).

5.3.1. *Convergence analysis*

To analyze the effect of the complete HOCT method, taking into account the modulation amplitude on the separation method, a convergence study is set up by varying the modulation amplitude of s_1 (from $f_A = 0$ Hz to $f_A = 70$ Hz) and the number of higher-order CTs computed. The $\mathrm{SNR}_{\mathrm{out}}$ of the estimated amplitude $\widetilde{A}_{s_1}$, with separation between the two components s_1 and s_3, is computed for each configuration. Results are shown in Fig. 9 (a). The highest $\mathrm{SNR}_{\mathrm{out}}$ for each value of the amplitude modulation metric (Eq. (32)) is shown on an optimum path. The relative error of the estimated amplitude for three cases $N = 1$, $N = 3$, and $N = 5$ is displayed Fig. 9 (b).

Compared to the results in Fig. 5 when dealing only with the amplitude modulation (no crossing ridges), the $\mathrm{SNR}_{\mathrm{out}}$ in Fig. 9 (a) clearly does not converge as the number N of higher-order moments increases. For a given amplitude modulation, the estimated signal amplitude rather reaches a maximum before gradually decreasing for higher N. Focusing on the details of the relative error of the estimated amplitude (see Fig. 9 (b): for $N = 1$), the order is too small to take into account the modulated amplitude. For $N = 3$ (optimum $\mathrm{SNR}_{\mathrm{out}}$), the modulated amplitude is better estimated and the separation is efficient even if not perfect ($\max_\tau \varepsilon = 2 \times 10^{-3}$). For $N = 5$, the estimate reaches a perfect reconstruction of the modulated amplitude (relative error lower than 1×10^{-4}) where the ridges are far from each other in the TF plane, but it has higher local relative error than for a smaller order $N = 3$ near the crossover. Trying to account for amplitude modulation with too many high-order chirplets degrades the separation of crossing components: compromise has to be made between separation quality and accuracy of the amplitude modulation estimate. The maximum $\mathrm{SNR}_{\mathrm{out}}$ decreases as the amplitude modulation metric increases, indicating that the separation method fails for very strong amplitude modulation.

6. Results on real signals

In this section, the proposed method is applied on real signals from vocal tract resonance measurements presented in Sec. 1 [31]. The record is a mixture of an amplitude modulated linear chirp which is the signal of interest (characterization signal), harmonics of the chirp coming from the non-linear behavior of the loudspeaker system, and voice signal (vowel /a/ around 440 Hz) considered as stationary (see Fig. 10 (a) left). It is sampled at 44100 Hz for 1 s. The instantaneous frequency of the characterization signal is perfectly known with just one delay, because vocal tract acoustics do not modify the instantaneous frequency: it is the same as the chirp emitted at the lips. Since the voice is louder than the chirp, and assumed to be stationary during a 1 s measurement record, the fundamental frequency is easily estimated as a local maximum of the spectrum. Frequency of voice harmonics are computed as multiple of the fundamental frequency. A discussion of the consequences of a bias in the knowledge of instantaneous component frequency, which is beyond the scope of this document, can be found in [31]. The HOCT method (Eq. (26)) is applied to extract the characterization signal from the mixture because the accuracy of the estimate of vocal tract resonances (out of scope of this paper) depends on the recovering of its amplitude modulation. The nearest voice harmonic from the local frame along the ridge is considered for being separated from the chirp, so we are dealing with only two components at the same time. The result is illustrated by the spectrogram ($N = 3$ higher-order moments), the waveform ($N = 0$ and $N = 3$) and the spectrum ($N = 0$ and $N = 3$) of the extracted signal on Fig. 10(a) right, (b) and (c), respectively.

The amplitude-modulated chirp is well-extracted from the record. Using higher-order moments ($N = 3$ compared to $N = 0$) from the HOCT method enables to better recover its amplitude modulation (see Fig. 10(b)) while voice harmonics are still separated (see Fig. 10(c)).

The computing time (1 s of signal sampled at 44100 Hz), on a Dell Precision 7530 (Debian) with CPU IntelCore i7-8750H (12 cores) and RAM 16GB, is 0.3s and 0.6 s, for $N = 0$ and $N = 3$ respectively. The method is therefore below real-time efficiency for real audio signals. Moreover, computation could be easily parallelized as the analysis is based on independent frames along the ridge, paving the way for application to high-dimensional data.

7. Conclusion

A complementary method to parametric time-frequency analyses, based on high-order chirplet transform (HOCT), has been presented to deal with amplitude-modulated signal reconstruction and components separation. In the framework of the vocal tract resonances measurement, with a fast frequency-varying, amplitude-modulated linear chirp superimposed on harmonics of the voice signal, we have focused on the quality of amplitude reconstruction assuming an *a priori* knowledge of the instantaneous frequency of each component. Relaxing the time-asymptotic hypothesis of the signal amplitude (i.e. constant am-

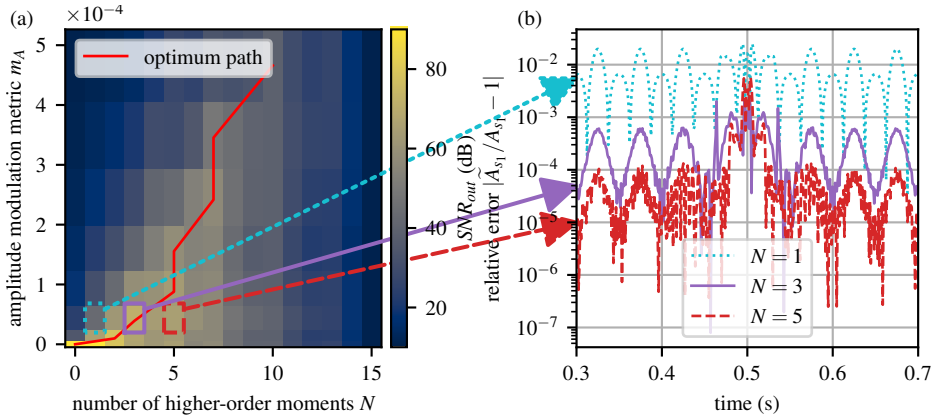

Figure 9: Convergence analysis : (a) optimum path as best SNR_{out} for a given amplitude-modulated signal s_1 separated from s_3 ; (b) zoom from 0.3 s to 0.7 s of the relative error of cases $N = 1$, $N = 3$ and $N = 5$ for s_1 (SNR_{out} = 42.8 dB, 70.8 dB and 64.6 dB, respectively).

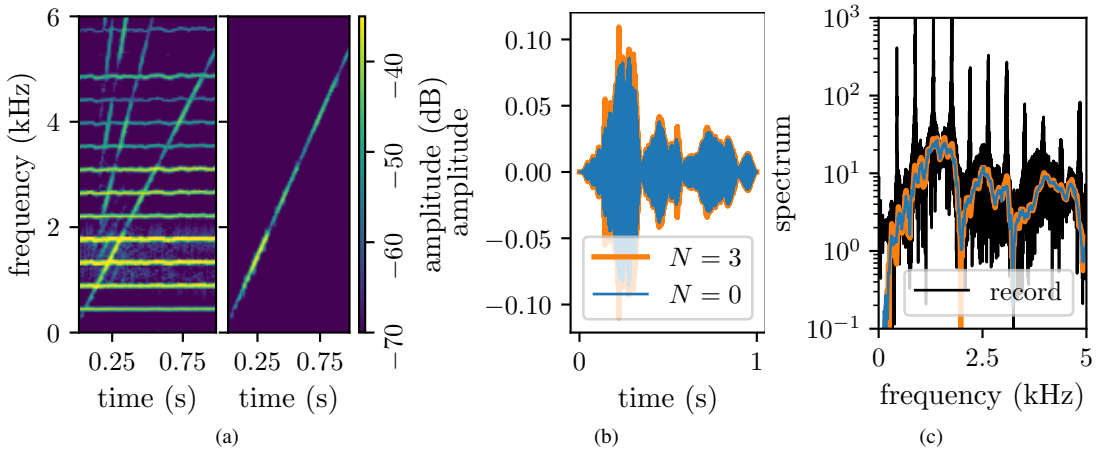

Figure 10: (a) Spectrograms of real record (left) and result of HOCT method separating the AM chirp with $N = 3$ higher-order moments (right) ; (b) waveform of separated chirp ($N = 0$ and $N = 3$) ; (c) amplitude spectrum of record and separated chirp ($N = 0$ and $N = 3$).

plitude within the window support), a set of chirplet windows based on Hermite polynomials has been built to ensure the numerical stability in the estimation of signal amplitude modulation. Introducing an amplitude modulation metric, the numerical results of convergence analysis show that an almost perfect reconstruction can be reached with a few higher-order CT moments, for most real signals presenting amplitude modulation. Constant amplitude crossing ridges in the TF plane are perfectly separated using adapted chirp rates. Finally, the HOCT method deals with amplitude-modulated crossing ridges and has been been illustrated in a simplified case with analytical solving. Numerical results show limitations in reconstruction because of a compromise between separation and accounting in amplitude modulation, leading to an optimal number of higher-order moments with respect to the SNR_{out}. The method is applied with real-time efficiency to real signals from vocal tract resonance measurements. Future studies should focus on the generalization of the method to deal with several jointly crossing ridges, and a better understanding of the compromise issue. Integrating the reconstruction method with existing instantaneous frequency estimators would make it possible to relax the need to know *a priori* the time-frequency trajectory of components.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Timothée Maison: Methodology, Investigation, Visualization, Formal analysis, Software, Writing – original draft. **Fabrice Silva**: Conceptualization, Supervision, Methodology, Investigation, Software, Writing – original draft, Writing – review & editing **Nathalie Henrich Bernardoni**: Supervision, Writing – review & editing. **Philippe Guillemain**: Supervision, Writing – review & editing.

Acknowledgments

This work was achieved in the framework of a PhD thesis supported by Aix-Marseille University, and the research was

partly funded by l'Agence Nationale de la Recherche (ANR), project RayoVox ANR-21-CE42-0017. For the purpose of open access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission. We thank the anonymous reviewers for their valuable and constructive comments.

Appendix A. Proof of Eq. (25)

Reminding $\zeta_{\Delta\beta} = 1/\sqrt{1 + j\sigma^2\Delta\beta}$, matrix $\mathbb{G}_{n,l}$ writes

$$\begin{aligned}\mathbb{G}_{n,l}(\Delta\omega,\Delta\beta) &= \frac{1}{\sigma\sqrt{2\pi n!}}\int_{\mathbb{R}} H_l\left(\frac{t}{\sigma}\right)H_n\left(\frac{t}{\sigma}\right)e^{-\frac{1}{2}\frac{t^2}{\sigma^2}(1+j\sigma^2\Delta\beta)}e^{-j\Delta\omega t}dt\\ &= \frac{\zeta}{\sqrt{2\pi n!}}\int_{\mathbb{R}} H_l\left(\zeta u\right)H_n\left(\zeta u\right)e^{-\frac{1}{2}u^2}e^{-j\Delta\omega\zeta\sigma u}du\\ &\quad\text{with variable change } t=\zeta\sigma u\end{aligned}$$

We consider the generative function associated to Hermite polynomials:

$$\exp\left(yz-\frac{1}{2}z^2\right) = \sum_{n\geq 0} H_n(y)\frac{z^n}{n!}$$

Introducing two abstracted variables r and s, we focus on this following expression:

$$\begin{aligned}&\sum_{n,l\geq 0}\mathbb{G}_{n,l}(\Delta\omega,\Delta\beta)\frac{r^n}{n!}\frac{s^l}{l!}\\ &= \frac{\zeta}{\sqrt{2\pi n!}}\int_{\mathbb{R}}\sum_{n,l\geq 0} H_l\left(\zeta u\right)H_n\left(\zeta u\right)e^{-\frac{1}{2}u^2}e^{-j\Delta\omega\zeta\sigma u}du\frac{r^n}{n!}\frac{s^l}{l!}\\ &= \frac{\zeta}{\sqrt{2\pi n!}}\int_{\mathbb{R}} e^{\zeta u s-\frac{1}{2}s^2}e^{\zeta u r-\frac{1}{2}r^2}e^{-\frac{1}{2}u^2}e^{-j\Delta\omega\zeta\sigma u}du\\ &= \frac{\zeta}{n!}e^{-\frac{1}{2}(1-\zeta^2)(s^2+r^2)+\zeta^2 sr}e^{-\frac{1}{2}(\Delta\omega\zeta\sigma)^2}e^{-j\Delta\omega\zeta^2\sigma(s+r)}\end{aligned}\tag{A.1}$$

So that $\mathbb{G}_{n,l}$ can be interpreted as the coefficient of the Taylor series as follows:

$$\mathbb{G}_{nl}(\Delta\omega,\Delta\beta) = \frac{\zeta}{n!}e^{-\frac{1}{2}(\Delta\omega\zeta\sigma)^2}\frac{\partial^{n+l}}{\partial r^n\partial s^l}\left[e^{\frac{1}{2}(1-\zeta^2)(s^2+r^2)+\zeta^2 sr-j\Delta\omega\zeta^2\sigma(s+r)}\right]_{\substack{r=0\\s=0}}$$

Development of the derivation leads to

$$\begin{aligned}\mathbb{G}_{n,l}(\Delta\omega,\Delta\beta) &= \frac{\zeta}{n!}e^{-\frac{1}{2}(\Delta\omega\zeta\sigma)^2}\sum_{p=0}^{min(l,n)}\frac{n!l!}{p!(n-p)!(l-p)!}(-1)^{n+l}\\ &\times\zeta^{2p}\left(\sqrt{1-\zeta^2}\right)^{n+l-2p}H_{n-p}\left(\frac{j\Delta\omega\sigma\zeta^2}{\sqrt{1-\zeta^2}}\right)H_{l-p}\left(\frac{j\Delta\omega\sigma\zeta^2}{\sqrt{1-\zeta^2}}\right)\end{aligned}\tag{A.2}$$

Finally, considering only even orders enables to retrieve Eq. (25)

Appendix B. Explicit expressions in denoising case

Here are presented explicit expressions of matrix terms in the simplified case of a signal of two components with a second asymptotic time component. We also choose $\omega_1 = \omega_2 = \varphi_1'(\tau)$, $\beta_1 = \varphi_1''(\tau)$ and $\beta_2 = -\varphi_1''(\tau)$, which is the case presented in the numerical results section. The system is written as:

$$\left[\begin{array}{c} L^{f_0}(\tau,\varphi_1'(\tau),\varphi_1''(\tau))\\ \vdots\\ L^{f_N}(\tau,\varphi_1'(\tau),\varphi_1''(\tau))\\ \hline L^{f_0}(\tau,\varphi_1'(\tau),-\varphi_1''(\tau))\end{array}\right] = \left[\begin{array}{c|c} \mathbb{B}_{n,k}^{(f)} & \mathbb{C}_n^{(f)}\\ \hline \mathbb{D}_k^{(f)} & \mathbb{E}^{(f)}\end{array}\right]\left[\begin{array}{c}\alpha_0^1 e^{j\varphi_1(\tau)}\\ \vdots\\ \alpha_N^1 e^{j\varphi_1(\tau)}\\ \hline \alpha_0^2 e^{j\varphi_2(\tau)}\end{array}\right]$$

with

$$\begin{aligned}\mathbb{B}_{0,k>0}^{(f)} &= (-1)^{k+1}(2k-1)!!\ \text{ and }\ \mathbb{B}_{n,n}^{(f)} = 1\\ \mathbb{C}_n^{(f)} &= \mathbb{G}_{2n,0}(\Delta\varphi',-\Delta\varphi'')\\ &= \frac{\zeta_{-\Delta\varphi''}}{(2n)!}\left(1-\zeta_{-\Delta\varphi''}^2\right)^n e^{-\frac{1}{2}\left(\sigma\Delta\varphi'\zeta_{-\Delta\varphi''}\right)^2}H_{2n}\left(\frac{j\sigma\Delta\varphi'\zeta_{-\Delta\varphi''}^2}{\sqrt{1-\zeta_{-\Delta\varphi''}^2}}\right)\\ \mathbb{D}_{k=0}^{(f)} &= \mathbb{G}_{0,0}(0,2\varphi_1'') = \zeta_{2\varphi_1''}\\ \mathbb{D}_{k>0}^{(f)} &= \mathbb{G}_{0,2k}(0,2\varphi_1'') - H_{2k}(0)\mathbb{G}_{0,0}(0,2\varphi_1'')\\ &= \zeta_{2\varphi_1''}(-1)^k(2k-1)!!\left[(1-\zeta_{2\varphi_1''}^2)^k-1\right]\\ \mathbb{E}^{(f)} &= \mathbb{G}_{0,0}(\Delta\varphi',-2\varphi_1''(\tau)+\Delta\varphi'') = \zeta_{-2\varphi_1''(\tau)+\Delta\varphi''}e^{\frac{1}{2}\left(\sigma\Delta\varphi'\zeta_{-2\varphi_1''(\tau)+\Delta\varphi''}\right)^2}\end{aligned}$$

where $\Delta\varphi' = \varphi_1'(\tau) - \varphi_2'(\tau)$ and $\Delta\varphi'' = \varphi_1''(\tau) - \varphi_2''(\tau)$.

References

[1] D. H. Whalen, W.-R. Chen, C. H. Shadle, S. A. Fulop, Formants are easy to measure; resonances, not so much: Lessons from Klatt (1986)a), The Journal of the Acoustical Society of America 152 (2) (2022) 933–941.

[2] C. H. Shadle, H. Nam, D. H. Whalen, Comparing measurement errors for formants in synthetic and natural vowels, The Journal of the Acoustical Society of America 139 (2) (2016) 713–727.

[3] M. Fleischer, A. Mainka, S. Kürbis, P. Birkholz, How to precisely measure the volume velocity transfer function of physical vocal tract models by external excitation, PLoS ONE 13 (3) (2018) e0193708.

[4] M. Kob, J. Thilaken, N. Henrich Bernardoni, Vocal-tract impedance at the mouth – from 1995 to today A tribute to Joe Wolfe and John Smith, in: Proc. 5th Stockholm Music Acoustic Conference (SMAC), To Be Published, Stockholm, Sweden, 2023.

[5] T. Maison, B. Allain, P. Hoyer, F. Silva, P. Guillemain, N. Henrich Bernardoni, Practical Guidelines for Implementing Vocal Tract Resonances Characterization with Excitation at the Lips, in: Proc. 12th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications, Florence, Italy, 2021.

[6] G. Gloth, M. Sinapius, Analysis of swept-sine runs during modal identification, Mechanical Systems and Signal Processing 18 (6) (2004) 1421–1441.

[7] T. Y. Hou, Z. Shi, Sparse time-frequency decomposition based on dictionary adaptation, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374 (2065) (2016) 20150192.

[8] S. Chen, Z. Peng, Y. Yang, X. Dong, W. Zhang, Intrinsic chirp component decomposition by using Fourier Series representation, Signal Processing 137 (2017) 319–327.

[9] L. Cohen, Time-Frequency Analysis, Prentice Hall Signal Processing Series, Prentice Hall PTR, Englewood Cliffs, N.J, 1995.

[10] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, H.-T. Wu, Time-Frequency Reassignment and Synchrosqueezing: An Overview, IEEE Signal Processing Magazine 30 (6) (2013) 32–41.

[11] I. Daubechies, J. Lu, H.-T. Wu, Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool, Applied and Computational Harmonic Analysis 30 (2) (2011) 243–261.

[12] G. Yu, M. Yu, C. Xu, Synchroextracting Transform, IEEE Transactions on Industrial Electronics 64 (10) (2017) 8042–8054.

[13] D. Fourer, F. Auger, K. Czarnecki, S. Meignen, P. Flandrin, Chirp Rate and Instantaneous Frequency Estimation: Application to Recursive Vertical Synchrosqueezing, IEEE Signal Processing Letters 24 (11) (2017) 1724–1728.
[14] N. A. Khan, S. Ali, A robust and efficient instantaneous frequency estimator of multi-component signals with intersecting time-frequency signatures, Signal Processing 177 (2020) 107728.
[15] T. Oberlin, S. Meignen, V. Perrier, Second-Order Synchrosqueezing Transform or Invertible Reassignment? Towards Ideal Time-Frequency Representations, IEEE Transactions on Signal Processing 63 (5) (2015) 1335–1344.
[16] D.-H. Pham, S. Meignen, High-Order Synchrosqueezing Transform for Multicomponent Signals Analysis—With an Application to Gravitational-Wave Signal, IEEE Transactions on Signal Processing 65 (12) (2017) 3168–3178.
[17] X. Chen, H. Chen, Y. Fang, Y. Hu, High-Order Synchroextracting Time–Frequency Analysis and Its Application in Seismic Hydrocarbon Reservoir Identification, IEEE Geoscience and Remote Sensing Letters 18 (11) (2021) 2011–2015.
[18] S. Lv, Y. Lv, R. Yuan, H. Li, High-order synchroextracting transform for characterizing signals with strong AM-FM features and its application in mechanical fault diagnosis, Mechanical Systems and Signal Processing 172 (2022) 108959.
[19] Y. Yang, Z. Peng, W. Zhang, G. Meng, Parameterised time-frequency analysis methods and their engineering applications: A review of recent advances, Mechanical Systems and Signal Processing 119 (2019) 182–221.
[20] S. Mann, S. Haykin, The chirplet transform: Physical considerations, IEEE Transactions on Signal Processing 43 (11) (1995) 2745–2761.
[21] Z. K. Peng, G. Meng, F. L. Chu, Z. Q. Lang, W. M. Zhang, Y. Yang, Polynomial Chirplet Transform With Application to Instantaneous Frequency Estimation, IEEE Transactions on Instrumentation and Measurement 60 (9) (2011) 3222–3229.
[22] X. Zhu, H. Yang, Z. Zhang, J. Gao, N. Liu, Frequency-chirprate reassignment, Digital Signal Processing 104 (2020) 102783.
[23] X. Zhu, Z. Zhang, Z. Li, J. Gao, X. Huang, G. Wen, Multiple squeezes from adaptive chirplet transform, Signal Processing 163 (2019) 26–40.
[24] X. Zhu, Z. Zhang, J. Gao, B. Li, Z. Li, X. Huang, G. Wen, Synchroextracting chirplet transform for accurate IF estimate and perfect signal reconstruction, Digital Signal Processing 93 (2019) 172–186.
[25] L. Li, N. Han, Q. Jiang, C. K. Chui, A chirplet transform-based mode retrieval method for multicomponent signals with crossover instantaneous frequencies, Digital Signal Processing 120 (2022) 103262.
[26] T. Wu, W. Zhang, B. Zhang, H. Luo, Synchro-reassigning generalized Chirplet basis transform for analyzing signals with crossover frequencies, Mechanical Systems and Signal Processing 195 (2023) 110313.
[27] Z. Chen, H.-T. Wu, Disentangling modes with crossover instantaneous frequencies by synchrosqueezed chirplet transforms, from theory to application, Applied and Computational Harmonic Analysis 62 (2023) 84–122.
[28] C. K. Chui, Q. Jiang, L. Li, J. Lu, Analysis of a direct separation method based on adaptive chirplet transform for signals with crossover instantaneous frequencies, Applied and Computational Harmonic Analysis 62 (2023) 24–40.
[29] H. Chen, X. Zhou, X. Chen, P. Lan, Y. Hu, A statistical frequency-chirprate extractor for mode retrieval with crossover instantaneous frequencies, Signal Processing 212 (2023) 109174.
[30] L. Xu, L. Chen, Z. Wang, W. Jiang, Z. Li, Multiple synchro-tuning chirplet transform, Digital Signal Processing 144 (2024) 104252.
[31] T. Maison, Towards the characterization of dynamical resonators: Measuring vocal tract resonances in singing, Ph.D. thesis, Aix-Marseille Université (Jul. 2023).
[32] P. Guillemain, R. Kronland-Martinet, Characterization of acoustic signals through continuous linear time-frequency representations, Proceedings of the IEEE 84 (4) (1996) 561–585.
[33] X. Zhu, B. Li, K. Yang, Z. Zhang, W. Li, Parameter analysis of chirplet transform and high-resolution time-frequency representation via chirplets combination, Signal Processing 205 (2023) 108824.