

Estimation of frequency modulations on wideband signals; applications to audio signal analysis

Harold Omer and Bruno Torr sani

Aix-Marseille Universit , CNRS, Centrale Marseille, LATP, UMR 7353, 13453 Marseille, France

Email: harold.omer@latp.univ-mrs.fr bruno.torresani@latp.univ-mrs.fr

Abstract—The problem of joint estimation of power spectrum and modulation from realizations of frequency modulated stationary wideband signals is considered. The study is motivated by some specific signal classes from which departures to stationarity can carry relevant information and has to be estimated.

The estimation procedure is based upon explicit modeling of the signal as a wideband stationary Gaussian signal, transformed by time-dependent, smooth frequency modulation. Under such assumptions, an approximate expression for the second order statistics of the transformed signal’s Gabor transform is obtained, which leads to an approximate maximum likelihood estimation procedure.

The proposed approach is validated on numerical simulations.

I. INTRODUCTION

Usual time-frequency models for audio signals often rest upon expansions with respect to dictionaries of time-frequency waveforms, such as Gabor frames, wavelet frames, or more general families. Such descriptions are generally adequate for signal classes such as (voiced) speech, music,... where specific time-frequency localisation properties can be exploited. They are less effective for less structured signals, such as wideband sound signals.

We are concerned here with an alternative description of audio signals, aiming at describing different sound classes such as environmental noise, engine sound,... which are in addition non-stationary, in the sense that they carry information related to dynamics. As an example, think of an accelerating engine sound, where the acceleration can generally be perceived. This example suggests to study sound models, which we will term *timbre* × *dynamics*, in which a reference (stochastic) stationary signal, characterized by its timbre, is modulated by some dynamic deformation. Given such signals, a problem is to estimate the modulation (and possibly the underlying power spectrum). While many techniques have been developed for frequency modulation estimation for narrow band signals (see e.g. [1], the wideband case is more complex and has apparently received less attention.

A class of models based upon deformations of stationary processes has been proposed and studied in [2], motivated by the famous *shape from texture* image processing problem. A main aspect of the approach is based on the remark that a generic class of transformations can be represented by transport equations in a suitable representation space.

We adopt here a more explicit point of view, and limit to stationary Gaussian processes, transformed by a time-dependent

modulation. We characterize the distribution of fixed time slices of a Gabor transform of such signals, and formulate the corresponding maximum likelihood estimation problem. As a result, we provide an estimation algorithm which is demonstrated on a small number of numerical examples.

II. FREQUENCY MODULATION OF STATIONARY RANDOM SIGNALS

A. Notations and background

1) *Random signals*: We shall be concerned with complex Gaussian random signal models \mathbf{X} of finite length L , which we shall assume zero-mean for the sake of simplicity. As is customary in finite-dimensional Gabor analysis, we shall also assume periodic boundary conditions, i.e. $\mathbf{X}_{t+L} = \mathbf{X}_t$. Given such a signal \mathbf{X} , we shall denote by $C_{\mathbf{X}}$ its covariance matrix, and by $R_{\mathbf{X}}$ its relation matrix (see [3] for details), defined as

$$C_{\mathbf{X}}(t, s) = \mathbb{E} \{ X_t \overline{X_s} \} , \quad R_{\mathbf{X}}(t, s) = \mathbb{E} \{ X_t X_s \} , \quad (1)$$

and we will write $\mathbf{X} \sim \mathcal{CN}(0, C_{\mathbf{X}}, R_{\mathbf{X}})$. \mathbf{X} is said to be *circular* if $R_{\mathbf{X}} = 0$.

2) *Time-frequency representation*: We shall use the following notations. Given a window function g , the corresponding short time Fourier transform of a signal (STFT) $\mathbf{x} \in \mathbb{C}^L$ is defined by

$$\mathcal{V}_g \mathbf{x}(m, n) = \sum_{t=0}^{L-1} x[t] \overline{g[t-n]} e^{-2i\pi m(t-n)/L} . \quad (2)$$

Given lattice constants a and b (divisors of the signal length L) the corresponding Gabor transform reads

$$\mathcal{G}_{\mathbf{x}}[m, n] = \mathcal{V}_g \mathbf{x}(mb, na) \quad m = 0, \dots, M-1, \quad n = 0 \dots N-1 , \quad (3)$$

with $M = L/b$ and $N = L/a$. $\mathcal{G}_{\mathbf{x}}$ is an $M \times N$ array. For suitably chosen g , and a and b small enough, the Gabor transform is invertible (see [4], [5]); in finite dimensional situations, efficient algorithms have been developed and implemented (see [6]).

Remark 1 (Notations): As usual, summation bounds in the frequency domain depend of the parity of the signal length L . For the sake of simplicity, we introduce some notations and denote by I_L the integer interval $I_L = [1 - L/2, L/2]$ if L is even, and the integer interval $I_L = [-(L-1)/2, (L-1)/2]$ if L is odd. The corresponding positive frequencies interval will be denoted by $I_L^+ = [0, L/2]$ if L is even, and $I_L^+ = [0, (L-1)/2]$ if L is odd.

B. The model: definition and main estimates

We are concerned here in a simple model of signal transformation, which may be written as follows. We denote by \mathbf{X} a zero-mean, wide sense stationary Gaussian random process, with covariance matrix $C_{\mathbf{X}}$, and by \mathbf{Z} the associated analytic signal. We denote by $\mathcal{S}_{\mathbf{X}}$ the power spectrum of \mathbf{X} , and assume that $\mathcal{S}_{\mathbf{X}}(0) = 0$, and if L is even, that $\mathcal{S}_{\mathbf{X}}(L/2) = 0$. Under such an assumption, it is easy to show that \mathbf{Z} is a circular complex Gaussian random vector (by a finite dimensional version of a standard argument, see e.g. [7]).

The observation is assumed to be the real part $Y_r = \text{Re}(\mathbf{Y})$ of a complex valued signal \mathbf{Y} ; for the sake of simplicity we shall only work with the latter, assumed to be an USB (upper sideband) modulated version \mathbf{Y} of a reference stationary signal \mathbf{X} , of the form

$$Y_t = Z_t e^{2i\pi\gamma(t)/L} + N_t, \quad (4)$$

where $\gamma \in C^2$ is an unknown smooth, slowly varying modulation function, and $\mathbf{N} = \{N_t, t = 0, \dots, L-1\}$ is a real Gaussian white noise, with variance σ_0^2 . Obviously, when γ is not a constant function, \mathbf{Y} is not a wide sense stationary signal any more. The problem at hand is to estimate the unknown modulation γ and the original power spectrum $\mathcal{S}_{\mathbf{X}}$ from a single realization of \mathbf{Y} .

Clearly, $\mathbf{Z} \sim \mathcal{CN}(0, C_{\mathbf{Z}}, 0)$ is a circular complex Gaussian random signal, with covariance matrix

$$C_{\mathbf{Z}}(t, s) = \sum_{\nu \in I_L^+} \mathcal{S}_{\mathbf{X}}(\nu) e^{2i\pi\nu(t-s)/L}, \quad (5)$$

and is therefore wide-sense stationary.

In the proposed approach, we will base the estimation on a Gabor representation of the observed signal, and deliberately disregard correlations across time of the Gabor transform (hence focusing on time slices of the Gabor transform of the observation). The distribution of time slices of the analytic signal \mathbf{Z} of the original signal is characterized in the following two results, which result from direct calculations.

Proposition 1: For fixed n , the Gabor transform $\mathcal{G}_{\mathbf{N}}[\cdot, n]$ of the gaussian white noise is a stationary Gaussian random vector, with circular covariance matrix

$$C_{\mathcal{G}_{\mathbf{N}}}[m, m'] = \sigma_0^2 \sum_{k=0}^{L-1} \bar{\hat{g}}[k] \hat{g}[k - (m' - m)b] \quad (6)$$

Proposition 2: For fixed time index n , the Gabor transform $\mathcal{G}_{\mathbf{Z}}[\cdot, n]$ of the analytic signal is a circular complex Gaussian random vector, with covariance matrix

$$C_{\mathcal{G}_{\mathbf{Z}}}[m, m'] = \sum_{k \in I_L^+} \mathcal{S}_{\mathbf{X}}[k] \bar{\hat{g}}[k - mb] \hat{g}[k - m'b] \quad (7)$$

The estimation of the modulation will be based upon an approximation of the covariance matrix of the observed signal. In a few words, the Gabor transform of the frequency modulated signal can be approximated by a deformed version of the Gabor transform of the original signal. The deformation

takes the form of a time-varying frequency shift. A more precise argument, based upon first order approximation of the modulation function γ , leads to the following result.

Theorem 1: 1) For fixed time, the Gabor transform $\mathcal{G}_{\mathbf{Y}}$ may be approximated as

$$\mathcal{G}_{\mathbf{Y}}[m, n] = \mathbf{G}^{(n; \gamma'(na)/b)}[m] + R[m], \quad (8)$$

where $\mathbf{G}^{(n; \delta)}$ is a frequency-shifted Gabor transform

$$\mathbf{G}^{(n; \delta)}[m] = \sum_{t=0}^{L-1} Z_t \bar{g}[t - na] e^{-2i\pi[m - \delta][t - an]/M} + \mathcal{G}_{\mathbf{N}}[m, n], \quad (9)$$

and the remainder is bounded as follows: for all m, m' ,

$$|\mathbb{E}\{\mathcal{G}_{\mathbf{Y}}[m] \overline{\mathcal{G}_{\mathbf{Y}}[m']}\}| \leq \sigma_Z^2 \left(\frac{\pi e}{L} \|\gamma''\|_{\infty} \mu_2 + 2\mu_1 \right)^2, \quad (10)$$

where σ_Z^2 is the variance of Z and with

$$\mu_1 = \sum_{t \in I_T^c} |g(t)|, \quad \mu_2 = \sum_{t \in I_T} t^2 |g(t)|, \quad T = \sqrt{\frac{L}{\pi \|\gamma''\|_{\infty}}} \quad (11)$$

where $I_T = [-T, T]$ and $I_T^c = I_L \setminus I_T$

2) Given δ , and for fixed n , $\mathbf{G}^{(n; \delta)}$ is distributed following a circular multivariate complex Gaussian law, with covariance matrix

$$C_{\mathbf{G}^{(n; \delta)}}[m, m'] = C_{\mathcal{G}_{\mathbf{Z}}}[m - \delta, m' - \delta] + C_{\mathcal{G}_{\mathbf{N}}}[m, m']. \quad (12)$$

The estimation procedure described below is a maximum likelihood approach, which requires inverting the covariance matrix of vectors $\mathbf{G}^{(n; \delta)}$. The latter is positive semi-definite by construction, but not necessarily definite. The result below provides a sufficient condition on \mathbf{g} and the noise for invertibility.

Proposition 3: Assume that the window \mathbf{g} is such that

$$K_{\mathbf{g}} := \min_{t=0 \dots L-1} \left(\sum_{k=0}^{b-1} |g[t + kM]|^2 \right) > 0. \quad (13)$$

Then for all $\mathbf{x} \in \mathbb{C}^M$,

$$\mathbf{x}^* C_{\mathbf{G}} \mathbf{x} \geq \sigma_0^2 K_{\mathbf{g}}, \quad (14)$$

and the covariance matrix is therefore boundedly invertible.

Remark 2: The condition may seem at first sight unnatural to Gabor frame experts. However, it simply expresses that the number M of frequency bins shouldn't be too large if one wants the covariance matrix to be invertible. However, reducing M also reduces the precision of the estimate, and a trade-off has to be found, as discussed in the next section.

C. Improving the frequency resolution

We propose here a method to improve the frequency resolution of our estimations. We have already seen that the invertibility of the covariance matrix requires that the number of frequency bins of the Gabor transform shouldn't be too large. As a result however, it may be convenient, as we shall

see later, to have access to the information contained in all the frequency frames of the short time Fourier transform defined in equation (2). For this purpose, we also consider alternative versions of the Gabor transform, associated with frequency-shifted sampling lattices:

$$\mathcal{G}_x^c[m, n] = \mathcal{V}_g \mathbf{x}(mb + c, na), \quad m \in \mathbb{Z}_M, n \in \mathbb{Z}_N, \quad (15)$$

where $c \in [0, b - 1]$. We now have at our disposal a collection of b Gabor transforms, which are all different subsampled versions of the STFT. The previous results and proofs remain valid with this new definition of the Gabor transform. Equations (8) and (9) now become

$$\mathcal{G}_Y^c[m, n] = \mathbf{G}^{(n; \gamma'(na)/b + c/b)}[m] + R, \quad (16)$$

where

$$\begin{aligned} \mathbf{G}^{(n; \delta^c)}[m] &= \sum_{t=0}^{L-1} Z_t \bar{g}[t - na] e^{-2i\pi[m - \delta^c][t - an]/M} \\ &\quad + \mathcal{G}_N^c[m, n], \end{aligned} \quad (17)$$

and the associated Equation (12) now reads:

$$C_{\mathbf{G}^{(n; \delta^c)}}[m, m'] = C_{\mathcal{G}_Z}[m - \delta^c, m' - \delta^c] + C_{\mathcal{G}_N}[m, m']. \quad (18)$$

The rationale will be that a frequency shift δ^c can be estimated from each one of these transforms, and the optimal one will be retained.

III. ESTIMATION PROCEDURE

We now describe in some details the estimation procedure corresponding to our problem. The estimation problem is the following: from a single realization of the signal model (4), estimate the modulation function γ and the original power spectrum \mathcal{S}_X . We first notice the indeterminacy in the problem, namely the fact that adding an affine function to γ is equivalent to shifting \mathcal{S}_X . This has to be fixed by adding an extra constraint in the estimation procedure.

A. Maximum likelihood modulation estimation

We now turn to the estimation procedure, that exploits the above results. With the same notations as before, we fix a value of the time index n , and denote for simplicity by $\mathcal{G} = \mathcal{G}^{(n)}$ the corresponding fixed time slice of \mathcal{G}_Z^c . Due to the multivariate complex Gaussian distribution of the signal and the fixed time Gabor transform slices, the log-likelihood of a slice takes the form

$$\mathcal{L}_\delta(\mathcal{G}) = \mathcal{G}^* (C_{\mathbf{G}^{(n; \delta^c)}})^{-1} \mathcal{G} + \ln(\pi^M \det(C_{\mathbf{G}^{(n; \delta^c)}})) . \quad (19)$$

Therefore, the maximum likelihood estimate for the frequency shift assumes the form

$$\hat{\delta}^c = \arg \min_{\delta^c} \left[\mathcal{G}^* (C_{\mathbf{G}^{(n; \delta^c)}})^{-1} \mathcal{G} + \ln(\pi^M \det(C_{\mathbf{G}^{(n; \delta^c)}})) \right] . \quad (20)$$

However, we notice that $\det(C_{\mathbf{G}^{(n; \delta^c)}})$ actually does not depend on the modulation parameter δ^c . Therefore the maximum likelihood estimate reduces to

$$\hat{\delta}^c = \arg \min_{\delta^c} \left[\mathcal{G}^* (C_{\mathbf{G}^{(n; \delta^c)}})^{-1} \mathcal{G} \right], \quad (21)$$

a problem to be solved numerically. Notice that this requires the knowledge of the covariance matrix $C_{\mathbf{G}^{(n; 0)}}$ corresponding to the Gabor transform of the noisy stationary signal. The latter is generally not available, and has to be estimated as well.

As $\delta^c(n) \approx (\gamma'(an) - c)/b$, the estimates of δ for each n lead to an estimate of γ' . Since we solve the minimisation problem by an exhaustive search on the δ^c , the estimate of γ' is coarsely quantized (see Remark 2), as b is large and $\hat{\gamma}'(an) \in [c, b + c, 2b + c, \dots, (M - 1)b + c]$. This problem is solved by using the family of frequency-shifted versions of Gabor transform described in subsection II-C and making a new exhaustive search on the $\hat{\delta}^c$

$$\hat{\delta} = \arg \min_c \left[\mathcal{G}^* (C_{\mathbf{G}^{(n; \delta^c)}})^{-1} \mathcal{G} \right] . \quad (22)$$

The quantization effect on the final estimation of the modulation function is therefore attenuated, i.e. $\hat{\gamma}'(an) \in [0, L - 1]$. Obtaining from this estimation a smoother estimate for the modulation function γ requires extra interpolation techniques.

Remark 3: As an alternative, one may also avoid exhaustive searches and seek minimizers in (21) using more elaborate numerical techniques, that would avoid quantization effects. This question is currently under study.

B. Estimation of the underlying covariance matrix

We now describe a method for estimating the covariance matrix $C_{\mathbf{G}^{(n; 0)}}$. Suppose that an estimate $\hat{\gamma}$ of the modulation function γ is available. Then the signal \mathbf{Y} can be demodulated by setting

$$\mathbf{U} = \mathbf{Y} e^{-2i\pi\hat{\gamma}/L}, \quad (23)$$

Clearly, \mathbf{U} is an estimator of $\mathbf{Z} + \mathbf{N} e^{-2i\pi\gamma/L}$, the noisy stationary signal. We can now compute the covariance matrix $C_{\mathcal{G}_U}$ of the Gabor transform of \mathbf{U} , which is an estimator of $C_{\mathcal{G}_Z} + C_{\mathcal{G}_N}$. Comparing with equation (9) we finally obtain an estimator for the covariance matrix

$$C_{\mathcal{G}_U} \approx C_{\mathbf{G}^{(n; 0)}} \quad (24)$$

Remark 4: The power spectrum \mathcal{S}_X of the stationary signal can be estimated from \mathbf{U} using a standard Welch periodogram estimator, or by marginalizing the square modulus of the Gabor transform of the demodulated signal, as described in [4].

C. Summary of the estimation procedure

We now summarize an iterative algorithm to jointly estimate the covariance matrix $C_{\mathbf{G}^{(n; 0)}}$ and the modulation function γ , that exploits alternatively the two procedures described above. The procedure is as follows, given a first estimation of the modulation function, we can perform a first estimation of the covariance matrix, which in turn allows us obtain a new estimation of the modulation function. The operation is repeated until the stopping criterion is satisfied.

For the initialization, we need a first modulation frequency estimate, for which we use the center of mass of the modulated signal Gabor transform

$$\hat{\delta}^{(0)}(n) = \frac{\sum_{m=0}^{M-1} m |\mathbf{G}^{(n; \delta)}|^2[m]}{\sum_{m=0}^{M-1} |\mathbf{G}^{(n; \delta)}|^2[m]} . \quad (25)$$

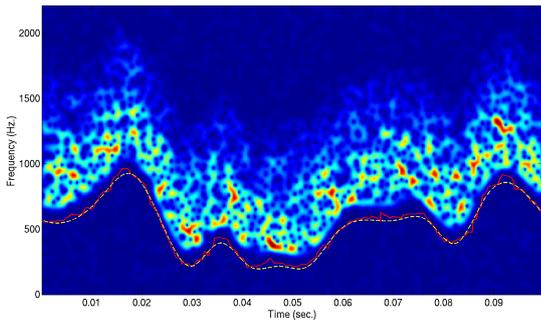


Fig. 1. Gabor transform of a frequency modulated synthetic stationary random signal, superimposed with the frequency modulation: estimate (red) and original (yellow).

The stopping criterion is based upon the evolution of the frequency modulation along the iterations. More precisely, we use the empirical criterion

$$\frac{\|\hat{\delta}^{(k)} - \hat{\delta}^{(k+1)}\|_2}{\|\hat{\delta}^{(k+1)}\|_2} < \epsilon \quad (26)$$

The pseudo-code of the algorithm can be found below

Algorithm 1 Joint covariance and modulation estimation

Initialize as in (25)

while criterion (26) is false **do**

- Compute $\hat{\gamma}^{(k)}$ by interpolation from $\delta^{(k)}$.
- Demodulate \mathbf{Y} using $\hat{\gamma}^{(k)}$ following (23)
- Compute the Gabor transform of the demodulated signal $\hat{\mathbf{G}}^{(k;n)}[m] = \mathcal{G}_{\mathbf{U}^{(k)}}[m, n]$
- Estimate $\hat{\delta}^{(k+1)}$ using the covariance matrix of $\hat{\mathbf{G}}^{(k;n)}$ from (21) and (22)
- $k := k + 1$

end while

IV. NUMERICAL RESULTS

The proposed estimation procedure has been implemented using MATLAB/OCTAVE, and relies on the LTFAT toolbox [8] for the time-frequency transforms.

We display in Fig. 1 an example of estimation result. The original signal was generated as pseudo-random stationary Gaussian signal with a smooth, wideband power spectrum, that was further modulated by a smooth frequency modulation function. Fig. 1 displays the Gabor transform of the modulated signal (positive frequencies only), together with the original and the estimate for the frequency modulation. For the sake of clarity, the frequency estimate has been displayed below the relevant part of the Gabor transform (remember that it is defined up to an additive constant). As can be seen, the result is fairly satisfactory, the estimated modulation follows closely the ground truth.

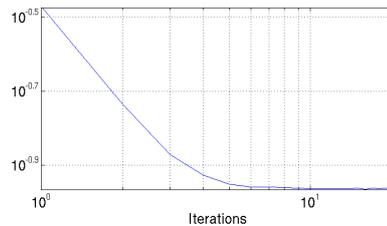


Fig. 2. Log-log plot of the evolution of the criterion proposed in (26) according on the number of iterations.

To assess the convergence properties of the proposed approach, the same experiment was run several times with the same modulation law and different seeds for the underlying stationary noise. We display in Fig. 2 the evolution of the criterion as a function of the iteration index, averaged over 20 realizations. Convergence appears to be fast, with power-law like decay speed.

V. CONCLUSION

We have presented in this paper a new approach for modulation frequency and power spectrum estimation from wideband signals, based upon explicit modeling. A main point that is exploited in our approach is the fact that modulations can be locally approximated by frequency shifts in the Gabor domain. The algorithm has been validated using numerical simulations, that show that when signals are generated according to the model of interest, very accurate results can be obtained.

Further developments include numerical tests on real signals, such as natural sounds generated by rolling bodies with variable speed,... We shall also consider extending this approach to other transformation models, such as time warping or more general transformations.

ACKNOWLEDGMENT

This work was supported by the ANR project Metason ANR-10-CORD-010.

REFERENCES

- [1] H. L. Van Trees, *Detection, estimation and modulation theory*. Wiley Interscience, 2003.
- [2] M. Clerc and S. Mallat, "Estimating deformations of stationary processes," *Annals of Statistics*, vol. 31, no. 6, pp. 1772–1821, 2003.
- [3] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2637–2640, 1996.
- [4] R. Carmona, W. L. Hwang, and B. Torr sani, *Practical time-frequency analysis: Gabor and Wavelet Transforms With an Implementation in S*, C. K. Chui, Ed. Academic Press, 1998.
- [5] K. Gr chenig, *Foundations of time-frequency analysis*, ser. Applied and Numerical Harmonic Analysis. Boston, MA: Birkh user Inc., 2001.
- [6] P. L. S ndergaard, "Efficient Algorithms for the Discrete Gabor Transform with a long FIR window," *J. Fourier Anal. Appl.*, vol. 18, no. 3, pp. 456–470, 2012.
- [7] B. Picinbono, "On circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.
- [8] P. S ndergaard, B. Torr sani, and P. Balazs, "The linear time frequency analysis toolbox," *International Journal of Wavelets and Multiresolution Information Processing*, vol. 10, no. 4, pp. 1 250 032–1 – 1 250 032–27, 2012.