# Informed Spectral Analysis for Isolated Audio Source Parameters Estimation

Dominique Fourer, Sylvain Marchand

# INFORMED SPECTRAL ANALYSIS FOR ISOLATED AUDIO SOURCE PARAMETERS ESTIMATION

*Dominique Fourer and Sylvain Marchand*

LaBRI CNRS
University of Bordeaux 1, 33405 Talence, France
firstname.name@labri.fr

## ABSTRACT

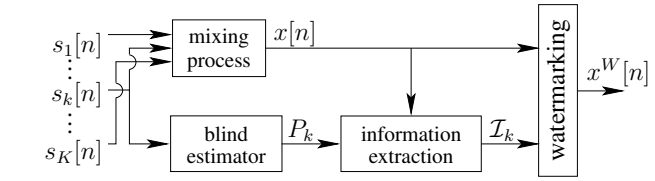
In this paper, we propose a new watermark-assisted method for the analysis of audio source signals present in a mixture. This work is motivated by the issue of quality-constrained source parameters estimation in under-determined mixtures where the blind approach is not efficient. Our method uses a specific coder-decoder configuration where the separated source signals are assumed to be known at the coder. The necessary information required by a classic blind estimator to reach a target quality is imperceptibly embedded into the mixture signal. At the decoder, where the isolated source signals are unknown, the analysis process is assisted by the extra information embedded into the mixture signal. Thus, this technique aims at opening new perspectives for quality-based audio applications like active listening, sound feature extraction, or high quality audio effects with a minimal amount of side information.

***Index Terms***— Informed spectral analysis, source separation, sinusoidal model

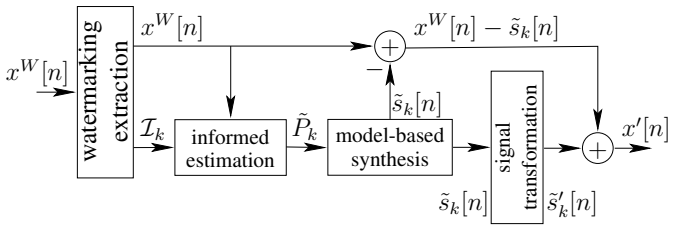
## 1. INTRODUCTION

Sound source separation is full of interest in audio processing and can allow a listener to manipulate each sound entity present in the auditory scene. The under-determined case (e.g. the number of sound entities is greater than the number of their observed mixtures) is a particularly difficult configuration. Moreover, it is the most frequent case for mono or stereo music mixtures. This issue is often processed using sparse representations of signals [1] but the quality is not sufficient for demanding applications. Recently, Informed Source Separation (ISS) [2] proposed to embed source indexes inaudibly into the mixture signal to assist the separation process. Thus, ISS achieves to reach a better quality of separation but does not use the full potential of available information present in the mixture. The method presented here is based on informed sinusoidal parameters estimation, and allows us to recover both the source signals and their model parameters, with a desired quality. The sinusoidal model offers a sparse representation of audio signals and is suitable for music. It has shown its efficiency for representing audio signals at low bit rates (typically lower than 24 kbits/s for MPEG-4 SSC [3]) and allows sound transformations like time stretching or pitch shifting with a high quality. Efficient estimator exists for the sinusoidal model [4] but have theoretical limitations on the best achievable estimation precision. As discussed in a previous work in [5], the unique way to break these theoretical bounds consists in injecting extra information into the analysis chain. With our approach, the information provided by a blind estimator is used to reduce the bit rate of this extra information. In the context where extra information is directly embedded into the signal mixture using a QIM-based watermarking method [6], differential coding is not applicable. In fact, the signal mixture is altered during the mixing and watermarking process and depends directly on the extra information itself. This paper proposes a solution to this issue and is organized as follows. Section 2 describes the principles of informed analysis, while its implementation is detailed in Section 3. Experiments and results for simulated and natural sounds are presented in Section 4. Section 5 concludes with discussions and future works.

## 2. INFORMED ANALYSIS FRAMEWORK




(a) Coder




(b) Decoder

Figure 1: Structure of the proposed system for informed single-source manipulation in under-determined sound mixture.

The presented method (see Fig. 1) is designed in a specific coder-decoder configuration. At the coder stage, a blind estimator is applied to each isolated source signal $s_k[n]$ to obtain reference parameters $P_k[n]$. After the mixing process, the same estimator is applied to the mixture signal $x[n]$ to obtain a blind estimation denoted $\hat{P}_k[n]$. The necessary information needed to recover $P_k[n]$ from $x[n]$ is estimated using the generalized informed spectral analysis technique detailed below. This section describes the signal model and the informed analysis principle applied to the parameters estimation at the coder and the decoder stages.

### 2.1. Sound source model and parameters estimation

Consider a discrete instantaneous single-channel mixture signal composed of $K$ sources which can be expressed as follows:

$$x[n] = \sum_{k=1}^{K} s_k[n] + r[n] \tag{1}$$

where $r[n]$ is the residual signal. Each source signal $s_k[n]$ is decomposed as a sum of sinusoids expressed as follows for a local analysis frame:

$$s_k[n] = \sum_{l=1}^{L} a_l \cos\left(\omega_l n + \phi_l\right) \tag{2}$$

where $a$, $\omega$, and $\phi$ correspond respectively to the amplitude, frequency, and phase parameters assumed to be locally constant. For the analysis process, instantaneous parameters are estimated using a classic frame-based estimator. An efficient estimator like spectral reassignment or the derivative method [4] is suitable for informed spectral analysis as discussed in [5]. In fact, these estimators achieve to reach the theoretical bounds and tend to reduce the bit rate required by informed spectral analysis.

### 2.2. Generalization of informed spectral analysis

Informed spectral analysis first introduced in [5] consists in assisting the analysis process using the minimal amount of extra information. This approach assumes that the initial exact parameters are available before the observed signal is altered during the mixing process. It consists in a two-step analysis where the information is first extracted using the knowledge about the expected value of the error resulting from a classic blind analysis. Second, the same estimator is applied on the altered signal and the errors are systematically corrected using the previously extracted information.

#### *2.2.1. Single parameter informed analysis*

Suppose we have to estimate a real parameter $0 \leq p < 1$ related to the signal $s_k[n]$. The mixture signal $x[n]$ is created according to (1) including the signal $s_k[n]$ combined with the others sources plus noise. The information needed to recover $p$ from a blind estimation $\hat{p}$ obtained from $x[n]$ is extracted as follows.

First, we define $\mathcal{C}_d : [0;1) \rightarrow \{0,1\}^d$ the $d$-bits precision fixed-point binary coding application and $\mathcal{D}$ the decoding application. $C = (C_1, C_2, ..., C_d)$ denotes the representation of $p$. Thus $\tilde{p}$ $d$-bits precision value of $p$ is decoded as follows:

$$\tilde{p} = \mathcal{D}(C) = \sum_{i=1}^{d} C_i 2^{-i}. \tag{3}$$

Second, $\mathcal{C}_d(p)$ is separated respectively in reliable part (the useful information provided by the blind estimator) and unreliable part (the less significant bits lower than the error standard deviation) as we have:

$$\mathcal{C}_d(p) = \underbrace{C_1, C_2, \ldots, C_{I_\sigma - 1}}_{\text{reliable part}}, \underbrace{C_{I_\sigma}, \ldots, C_d}_{\text{unreliable part}} \tag{4}$$

where $I_\sigma = \text{msb}(\mathcal{C}_d(\sigma))$ is the index of the most significant bit of $\mathcal{C}_d(\sigma)$, $\sigma$ denoting the standard deviation of the error. According to the informed spectral analysis principle [5], it is possible to recover $\tilde{p}$ ($d$-bits precision value of $p$) from any $\hat{p}$ if $|p - \hat{p}| \leq \sigma$ using the extra information:

$$\mathcal{I} = C_{I_\sigma - 1}, C_{I_\sigma}, \ldots, C_d. \tag{5}$$

At the decoder, a new $\hat{p}$ value is obtained from the analysis of $x[n]$. As explained in the introduction, $\hat{p}$ is unknown at the decoder due to the watermarking process dependent on the extra information embedded. Thus, estimation errors are systematically corrected by substitution of the unreliable part with $\mathcal{I}$ with the knowledge about $I_\sigma$ using the algorithm proposed in [5] to solve eventual carry or exception problems.

#### *2.2.2. Entire sinusoidal model informed analysis*

Consider now we have to estimate $P = (a, \omega, \phi)$ a vector of $\mathbb{R}^3$. As $a$, $\omega$, and $\phi$ have different physical meaning, they require a different relative accuracy in order to minimize a defined distortion measure. The distortion considered here is the quadratic error between synthesized signals according to (2). Thus, the function $b_{a,\omega,\phi}(m)$ which returns the number of bits allocated to each parameter for a maximal overall bit budget $m$ is computed. The bit allocation problem can be solved for an arbitrary probability distribution over $(a, \omega, \phi)$ using vector quantization [7]. However, the Entropy Constrained Unrestricted Spherical Quantization (ECUSQ) method described in [8] is preferred to the iterative approach. ECUSQ reduces the computation cost for a large set of data by the derivation of an analytic form for the optimal quantizer. The resulting bit rate is strongly reduced compared to classic scalar quantizers. The bit budget allocated to each parameter $P = (a, \omega, \phi)$ depends on the value of each component and results a variable bit rate for a fixed entropy $H_t$ (e.g. if $a \approx 0$, we need to allocate bits neither to phase nor to frequency). The entropy $H_t$ is deduced from the desired accuracy from distortion rate function defined in [8] and allows to define the effective bit budget $d$ and $b_{a,\omega,\phi}$. The extra information is extracted separately for each component of $P$ as it is explained for the single parameter case. Finally, $P$ is coded using a simple concatenation:

$$\mathcal{C}_d(P) = \mathcal{C}_e(a), \mathcal{C}_f(\omega), \mathcal{C}_g(\phi) \text{ with } e + f + g = d \tag{6}$$

where $e = b_a(d)$, $f = b_\omega(d)$, and $g = b_\phi(d)$. For the decoding process, the prior knowledge about $d$ and $I_\sigma$ of each parameter is necessary. With the ECUSQ method, $f$ and $g$ can only be computed from $\tilde{a}$ after the error correction. This point is detailed in the next section.

## 3. IMPLEMENTATION

### 3.1. Overall algorithm

The complete method presented in Fig. 1 is implemented for the coder and decoder according to Algorithms 1 and 2.

$I_{\sigma,k,l}$ is computed for each parameter component of the source $s_k[n]$. It can easily be estimated from the initial Signal-to-Noise Ratio (SNR) computed from $x[n]$ and $s_k[n]$ with the knowledge about the Cramèr-Rao Bounds (see [4]).

### 3.2. Quantization and coding

The ECUSQ method [8] provides analytic expressions for the sinusoidal parameters optimal quantizer with the high-resolution assumption (the error is uniformly distributed on each quantization

**Algorithm 1** Coder

**input:** $s_k[n]$: isolated source signals
**output:** $x^W[n]$: watermarked mixture

- Estimate $P_{k,l}[n]$ from $s_k[n]$ using reassignment method [4].
- Compute $b_{a,\omega,\phi}$ from $P_{k,l}[n]$ using the ECUSQ method [8].
- Compute binary mask$[n]$ containing the support of detected peaks in discrete amplitude spectrum.
- Estimate $\hat{P}_{k,l}[n]$ from random generated mixture combined with a random watermark computed according to (1) and method [9] for the watermark.
- Estimate $I_{\sigma,k,l}$ and $\mathcal{I}_{k,l}$ from $\hat{P}_{k,l}[n]$ using the informed spectral analysis method (see Section 2).
- Compute $x^W[n]$ using [9] containing mask$[n]$, $I_{\sigma,k,l}$ and $\mathcal{I}_{k,l}$.

**Algorithm 2** Decoder

**input:** $x^W[n]$ : watermarked mixture
**output:** $\tilde{s}_k[n]$, $\tilde{P}_{k,l}[n]$: isolated source signals and parameters

- Recover mask$[n]$, $I_{\sigma,k,l}$ and $\mathcal{I}_{k,l}$ from watermark extraction from $x^W[n]$ using [9].
- Estimate $\hat{P}_{k,l}[n]$ using mask$[n]$ and reassignment method [4].
- Compute $\tilde{P}_{k,l}[n]$ with $I_{\sigma,k,l}$ and $\mathcal{I}_{k,l}$ using the informed spectral analysis (see Section 2).
- Synthesize $\tilde{s}_k[n]$ from $\tilde{P}_{k,l}[n]$ according to (2).

cell). Thus for a target entropy (with a variable bit rate) related to a fixed resolution (constant bit rate), the average distortion measure is minimized. The unrestricted term refers to the fact that parameters are dependently quantized which outperform restricted spherical quantizers. The ECUSQ quantizer point density function are expressed as follows:

$$g_a(a,\omega,\phi) = 2^{(1/3)\tilde{H}_t - 2\beta(A) - \log_2(\sigma_w)}, \quad (7)$$

$$g_\phi(a,\omega,\phi) = a g_a(a,\omega,\phi), \quad (8)$$

$$g_\omega(a,\omega,\phi) = \sigma_w a g_a(a,\omega,\phi), \quad (9)$$

with $\tilde{H}_t = H_t - h(A)h(\Omega)h(\Phi)$ where $H_t$ is the target entropy and $h(.)$ denotes the entropy of each parameter probability distribution respectively denoted $f_A(a)$, $f_\Omega(\omega)$, and $f_\Phi(\phi)$. We define $\beta(A) = \int f_A(a)\log_2(a)da$ and $\sigma_w^2 = \frac{1}{||w||^2}\sum_{n=n0}^{n_0+N-1} w(n)n^2$. Here, $w(n)$ denotes the analysis window of size $N$. The quantization step sizes are given by the reciprocal values of the point densities $\Delta = g^{-1}$. Each quantization point is located in the middle of the corresponding quantization cell. The first quantization point is chosen to be 0. $g_\omega$ and $g_\phi$ depend linearly on the amplitude and are computed using $\tilde{a}$. T he number of bits allocated to each parameter for fixed-point binary coding is 0 if $g = 0$ and $\lceil \log_2(g) \rceil$ elsewhere.

### 3.3. Watermarking process

The extra information is inaudibly embedded into the mixture using the method described in [9]. This technique inspired from the Quantization Index Modulation (QIM) [6] is based on Modified Discrete Cosine Transform (MDCT) coefficients quantization. We selected this method for the large capacity provided, higher than 200kbits/s for 16-bit PCM signals and for its high resulting quality. The embedded extra information is designed for the frame-based spectral analysis at the decoder stage. It is composed of a binary mask of Fourier bins defined for each analyzed frame, in order to assist the peak extraction. The accuracy enhancement for each estimated parameter is provided by $\mathcal{I}_{k,l}$ and the $I_{\sigma,k,l}$ index concatenated for each sinusoidal component.

## 4. EXPERIMENTAL RESULTS

### 4.1. Simulation

Consider here a discrete-time signal $s[n]$ sampled at $F_s = 44.1$kHz composed of 1 sinusoid generated according to (2). A white Gaussian noise of fixed variance is mixed with $s[n]$ in order to result a SNR from $-20$dB to 50dB. For each SNR, 10000 random signals are generated. Phase follows the uniform density function $U(0, 2\pi)$. Amplitude and frequency follow a Rayleigh distribution respectively of parameter $\sigma_a = 0.2$ for $a \in [0, 1]$ and $\sigma_\omega = \pi/11$ for $\omega \in [0, \pi)$. The analysis frame uses the Hann window of odd length $N = 1023$ with estimation time set at the center. A target SNR is set at 45dB and the corresponding target entropy is deduced from the minimal ECUSQ distortion function [8]. For each generated signal, $I_\sigma$ is estimated using the knowledge about the SNR value uniformly vector quantized with 4 bits on the $[-20\text{dB}, \text{SNR}^{\text{target}}]$ interval.

### 4.2. Application to real sounds

For this experiment[1], a single-channel 44.1kHz-sampled music signal is processed. This musical piece is composed of a male singer voice and a rhythmic guitar. During the preliminary analysis step, sinusoidal parameters of the target source $s_k[n]$ signal are first blind estimated [4] with a Hann analysis window of length $N = 1023$ with 50% overlap. The instantaneous mixture signal is computed according to (1) with $s_k[n]$ synthesized from the reference parameters. Fig. 3(a) and 3(b) show the bit rate necessary to reach a target SNR respectively for the voice and the guitar source signals.

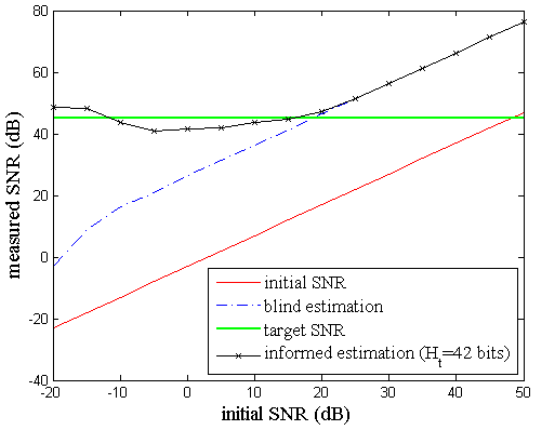
## 5. CONCLUSION AND FUTURE WORK

We have proposed a complete application for informed spectral analysis based on sinusoidal modeling combined with a watermarking technique. As shown on Fig. 3, the informed approach combined with the reassignment estimator achieves to reduce the bit rate of about 50% for the guitar signal and about 60% for the voice signal in comparison to uniformly quantized parameters. This method based on parameters coding can easily be adapted to others estimators and watermarking techniques to reach a fixed target quality. This method has still to be optimized using entropy coding to reduce the bit rate of the extra information.

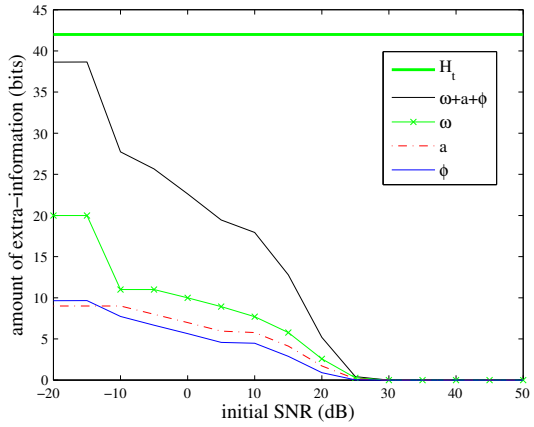
## 6. ACKNOWLEDGMENTS

This research was partly supported by the French ANR DReaM project (ANR-09-CORD-006).

[1] Sounds examples are available on-line at URL:
`http://www.labri.fr/~fourer/publi/WASPAA11/`

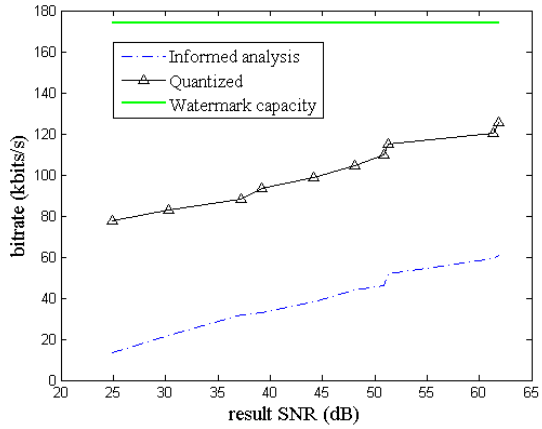



(a)

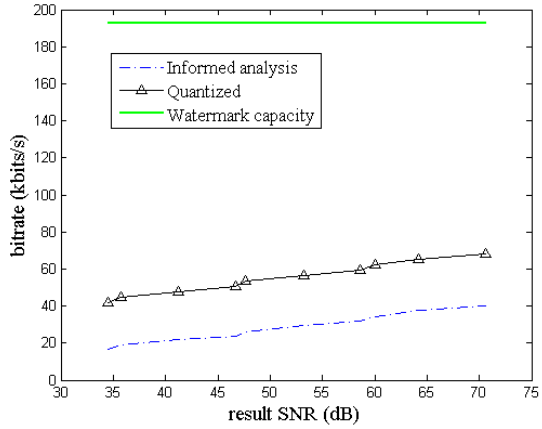



(b)

Figure 2: Average SNR 2(a) and corresponding average bit rate 2(b) for a target quality set at 45dB for 10000 random generated signals. The target entropy $H_t$ is deduced from ECUSQ distortion-rate function [8].




(a) Voice




(b) Guitar

Figure 3: Bit rate comparison between the full informed approach and informed analysis with 5-bit quantized $I_\sigma$ and respective $I_\sigma$ for each parameter.

## 7. REFERENCES

[1] P. Bofill and M.Zibulevski, "Underdetermined blind source separation," in *Signal Processing*, vol. 81, no. 11, 2001, pp. 2353–2362.

[2] M. Parvaix and L. Girin, "Informed source separation of underdetermined instantanenous stereo mixtures using source index embedding," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 245–248.

[3] E. Schuijers, W. Oomen, B. Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," in *Proc. 114th Conv. Audio Eng. Soc.*, Mar. 2003, pp. 201–204.

[4] S. Marchand and P. Depalle, "Generalization of the derivative analysis method to non-stationary sinusoidal modeling," in *Proc. DAFx Conf.*, Sep. 2008, pp. 281–288.

[5] S. Marchand and D. Fourer, "Breaking the bounds: Introducing informed spectral analysis," in *Proc. DAFx Conf.*, Sep. 2010, pp. 359–366.

[6] B. Chen and G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[7] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. on Acous., Speech, and Sig. Proc.*, vol. 37, no. 1, pp. 31–42, 1989.

[8] P. Korten, J. Jensen, and R. Heusdens, "High-resolution spherical quantization of sinusoidal parameters," *IEEE Trans. on Audio, Speech, and Lang. Proc.*, vol. 15, no. 3, pp. 966–981, Mar. 2007.

[9] J. Pinel, L. Girin, C. Baras, and M. Parvaix, "A high-capacity watermarking technique for audio signals based on MDCT-domain quantization," in *Int. Congress on Acoustics*, Oct. 2010.