ESPRIT in Gabor Frames

Adrien Sirdey, Olivier Derrien, Richard Kronland-Martinet, Mitsuko Aramaki

ESPRIT in Gabor frames

Adrien Sirdey[1], Olivier Derrien[1], Richard Kronland-Martinet[1], and Mitsuko Aramaki[1]

[1]*Laboratoire de Mécanique et d'Acoustique, CNRS, Marseille, France*

Correspondence should be addressed to Adrien Sirdey (sirdey<at>lma.cnrs-mrs.fr)

ABSTRACT

This articles tackles the estimation of mode parameters in recorded sounds of resonant objects. High resolution methods such as the ESPRIT method have already proved to be of great use for this sort of purpose. However, these methods being model-sensitive, their application to real-life audio signals can lead to results that are not satisfactory enough for a consistent re-synthesis. This is especially the case when the computational cost makes it impossible to analyse the signal in totality, or when the signal presents a high number of components. Significant improvements have already been achieved by decomposing the signal into several sub-band filtered versions, and by applying the ESPRIT algorithm on each of the resulting signals. It is shown in this article that the ESPRIT algorithm can be efficiently applied on time-frequency representations of the signal obtained using Gabor frames. Numerical tests that highlight the advantages of such an approach are also detailed. In addition to the advantages offered by the sub-band approach, the solid Gabor frame formalism combined with the ESPRIT method allows a flexible and sharp analysis on selected regions of the time-frequency plane, and leads to re-synthesis which are perceptually very close to the original sounds.

1. INTRODUCTION

The context of this study is the identification of acoustical modes which characterize a resonant object. This is of great use when building an environmental sound synthesizer (see [1] or [2] for an insight on such synthesizers). Practically, the analysis is made from recorded impact sounds, where the resonant object is hit by another solid object (e.g. a hammer). Assuming that the impact sound is approximately the acoustical impulse response of the resonant object, each mode corresponds to an exponentially damped sinusoid (EDS). The modal analysis thus consists of estimating the parameters of each sinusoidal component (amplitude, phase, frequency and damping). These parameters will be stored, and eventually modified, before further re-synthesis. In this paper, only the analysis part will be considered.

In the past decades, significant advances have been made in the field of system identification, especially for estimating EDS parameters in a background noise. Although the so-called *high-resolution methods* or *subspace methods* (MUSIC, ESPRIT) [3, 4] were proved to be more efficient than spectral peak-picking and iterative analysis-by-synthesis methods [5], few applications have been proposed. One can suppose that the high computational complexity of these methods is a major drawback to their wide use: on a standard modern computer, high-resolution methods can hardly analyse more than 10^4 samples, which corresponds roughly to 200 ms sampled at 44100 Hz. This is usually too short for analysing properly impact sounds which can last up to 10 s. Sub-band decomposition with critical sub-sampling in each band seems to be a natural solution to overcome the complexity problem, as it has already been shown in [6] and [7]. This can also be combined with a prior decomposition of the original signal in the time domain as shown in [8]. Another drawback is that ESPRIT gives accurate estimates when the background noise is white, which is usually not the case in practical situations. This problem can be overcome by the use of whitening filters. The estimation of the model order (i.e. the number of modes) is also an important issue. Various methods have been proposed for automatic estimation of the order, e.g. ESTER [9], but this parameter is often deliberately over-estimated in most practical situation.

In this paper, a novel method is proposed for estimating the modes with the ESPRIT algorithm, by applying it on a time-frequency representation of the original sound. The time-frequency representation is here computed within a *Gabor frame*, which forms a discrete paving of the time-frequency plane. The transform applied to the signal in order to express it in a given Gabor frame is called a *Gabor transform* (GT). Computing the

GT, a straightforward time subsampling and sub-band division of the signal is achieved. It is shown that an EDS in the original sound is still an EDS inside each frequency channel, and that ESPRIT can be applied in each of these channels in order to recover the original parameters. Furthermore, if the number of frequency sub-bands is high enough, it is reasonable to assume that the noise is white inside each sub-band, which renders the analysed samples more conform to the underlying mathematical model. A method to discard insignificant modes *a posteriori* is also proposed.

The paper is organised as follows: first, a brief state-of-the-art covers the signal model, the ESPRIT algorithm and the Gabor transform. Then, it is shown that original EDS parameters can be recovered by applying the ESPRIT algorithm in each frequency channel of the Gabor transform. The next part describes numerical tests that have been conducted in order to test the robustness of the method. Then, an experimentation on a real metal sound is described, and shows the efficiency of the proposed method. Further improvements are finally discussed.

2. STATE OF THE ART

2.1. The signal model and the ESPRIT algorithm

The discrete signal to be analysed is written:

$$x[l] = s[l] + w[l] \tag{1}$$

where the deterministic part $s[l]$ is a sum of K damped sinusoids:

$$s[l] = \sum_{k=0}^{K-1} \alpha_k z_k^l \tag{2}$$

where the complex amplitudes are defined as $\alpha_k = a_k e^{i\phi_k}$ (containing the initial amplitude a_k and the phase ϕ_k), and the poles are defined as $z_k = e^{-d_k + 2i\pi\nu_k}$ (containing the damping d_k and the frequency ν_k). The stochastic part $w[l]$ is a gaussian white noise of variance σ^2.

The ESPRIT algorithm was originally described by Roy *et. al.* [4], but many improvements have been proposed. Here, the Total Least Square method by Van Huffel *et. al* [10] will be used. The principle consists of performing a SVD on an estimate of the signal correlation matrix. The eigenvectors corresponding to the K highest eigenvalues correspond to the so called *signal subspace*, while the remaining vectors correspond to the so called *noise subspace*. The shift invariance property of the signal subspace allows a simple solution for the optimal poles values z_k. Then, the amplitudes α_k can be recovered by solving a least square problem. The algorithm can be described briefly as follows:

The signal vector is defined as:

$$\boldsymbol{x} = \left[\begin{array}{cccc} x[0] & x[1] & \dots & x[L-1] \end{array} \right]^T, \tag{3}$$

where L is the length of the signal to be analysed. The Hankel signal matrix is defined as:

$$\boldsymbol{X} = \left[\begin{array}{cccc} x[0] & x[1] & \dots & x[Q-1] \\ x[1] & x[2] & \dots & x[Q] \\ \vdots & \vdots & & \vdots \\ x[R-1] & x[R] & \dots & x[L-1], \end{array} \right] \tag{4}$$

where $Q, R > K$ and $Q + R - 1 = L$. The amplitude vector is defined as:

$$\boldsymbol{\alpha} = \left[\begin{array}{cccc} \alpha_0 & \alpha_1 & \dots & \alpha_{K-1} \end{array} \right]^T, \tag{5}$$

and the Vandermonde matrix of the poles:

$$\boldsymbol{Z}^L = \left[\begin{array}{cccc} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \cdots & z_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ z_0^{L-1} & z_1^{L-1} & \cdots & z_{K-1}^{L-1} \end{array} \right]. \tag{6}$$

Performing a SVD on $\boldsymbol{X}$ leads to:

$$\boldsymbol{X} = [\boldsymbol{U_1 U_2}] \left[\begin{array}{cc} \boldsymbol{\Sigma}_1 & 0 \\ 0 & \boldsymbol{\Sigma}_2 \end{array} \right] \left[\begin{array}{c} \boldsymbol{V_1} \\ \boldsymbol{V_2} \end{array} \right], \tag{7}$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are diagonal matrix containing respectively the K largest singular values, and the smallest singular values; $[\boldsymbol{U_1 U_2}]$ and $[\boldsymbol{V_1 V_2}]$ are respectively the corresponding left and right singular vectors. The shift-invariance property of the signal space leads to:

$$\boldsymbol{U}_1^{\downarrow} \boldsymbol{\Phi}_1 = \boldsymbol{U}_1^{\uparrow}, \qquad \boldsymbol{V}_1^{\downarrow} \boldsymbol{\Phi}_2 = \boldsymbol{V}_1^{\uparrow}, \tag{8}$$

where the eigenvalues of $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ provide an estimation of the poles z_k. $(.)^{\uparrow}$ and $(.)^{\downarrow}$ respectively stand for the operators discarding the first line and the last line of a matrix. Thus, z_k can be estimated by diagonalization of matrix $\boldsymbol{\Phi}_1$ or $\boldsymbol{\Phi}_2$. The associated Vandermonde matrix $\boldsymbol{Z}^L$ is computed. Finally, the optimal amplitudes with respect to the least square criterion are obtained by:

$$\boldsymbol{\alpha} = (\boldsymbol{Z}^L)^{\dagger} \boldsymbol{x}, \tag{9}$$

where $(.)^{\dagger}$ denotes the pseudoinverse operator.

2.2. The Gabor Transform

The Gabor transform allows the expression of $x[l]$ in a given Gabor frame. A Gabor frame $\{g,a,M\}$ is characterised by a window g, a time-step parameter a, and a number of frequency channels M. The expression $\chi[m,n]$ of $x[l]$ in the Gabor frame $\{g,a,M\}$ is written:

$$\chi[m,n] = \sum_{l=0}^{L-1} \overline{g}[l-an]x[l]\, e^{-2i\pi l \frac{m}{M}}, \tag{10}$$

where $\overline{(.)}$ denotes the complex conjugate. m is a discrete frequency index and n a discrete time-index. One can see that this corresponds to a discretised version of the standard short-time Fourier transform. For some frames, this transform can be inverted (for more details, see for instance [14]). The signal $\chi[m,n]$ for a fixed index m can be seen as a sub-sampled and band-pass filtered version of the signal $x[l]$. As the sub-sampling reduces the length of the data by a factor a, the ESPRIT algorithm can be applied to each frequency channel in order to analyse longer signals.

3. ESPRIT IN A GABOR FRAME

This section covers the application of the ESPRIT algorithm to a single channel in a Gabor frame. The analysed signals are therefore composed of the GT coefficients at a given frequency index m. As the GT is linear, the contribution of the deterministic part $s[l]$ can be separated from the contribution of the noise $w[l]$.

3.1. Deterministic part

$c[m,n]$ denotes the GT of $s[l]$ in channel m and time index n, whereas $c_k[m,n]$ denotes the GT of the signal z_k^l associated to the pole z_k:

$$c_k[m,n] = \sum_{l=0}^{L-1} \overline{g}[l-an] z_k^l\, e^{-2i\pi l \frac{m}{M}}. \tag{11}$$

According to the signal model (2), is can be easily proved that:

$$c[m,n] = \sum_{k=0}^{K-1} \tilde{\alpha}_{k,m} \tilde{z}_{k,m}^n, \tag{12}$$

where the apparent pole $\tilde{z}_{k,m}$ can be written as:

$$\tilde{z}_{k,m} = z_k^a\, e^{-2i\pi a \frac{m}{M}}, \tag{13}$$

and the apparent amplitude:

$$\tilde{\alpha}_{k,m} = \alpha_k\, c_k[m,0]. \tag{14}$$

In other words, the deterministic part of the signal in each channel is still a sum of exponentially damped sinusoids. However, their poles and amplitudes are modified according to the Gabor frame time-step and frequency parameters.

3.2. Stochastic part

Assuming that the time-step a is close to M ensures that the GT of the noise in each channel is approximately white. Furthermore, it has been proved that the Gabor transform of a gaussian noise is a complex gaussian noise [11]. It his thereafter assumed that the GT of $w[l]$ in each channel is a complex white gaussian noise.

3.3. Recovering the signal parameters

As the signal model is still valid, it is reasonable to apply ESPRIT on $c[m,n]$. $\boldsymbol{c}_m$ denotes the vector of GT coefficients in the channel m and $\boldsymbol{S}_m$ the Hankel matrix built from $c[m,n]$. Applying the ESPRIT algorithm to $\boldsymbol{S}_m$ leads to the estimation of the apparent poles $\tilde{z}_{k,m}$. Inverting equation (13) leads to:

$$z_k = e^{2i\pi \frac{m}{M}} (\tilde{z}_{k,m})^{\frac{1}{a}}. \tag{15}$$

Because of the sub-sampling introduced by the GT, it can be seen from equation (13) that aliasing will occur when the frequency of a pole is outside the interval $\left[\frac{m}{M} - \frac{1}{2a}, \frac{m}{M} + \frac{1}{2a}\right]$. To avoid aliasing, the analysis window $g[l]$ is chosen so that its bandwidth is smaller than $\frac{1}{a}$. That way, the possible aliasing components will be attenuated by the band-pass effect of the Gabor transform.

Denoting $\tilde{\boldsymbol{Z}}_m^N$ the Vandermonde matrix of the apparent poles $\tilde{z}_{k,m}$ (N is the time-length of signal $c[m,n]$), the least square method for estimating the amplitudes leads to:

$$\boldsymbol{\alpha} = \frac{(\tilde{\boldsymbol{Z}}_m^N)^{\dagger} \boldsymbol{c}_m}{c_k[m,0]}. \tag{16}$$

Without noise, according to equation (12), each EDS should be detected in each channel, which generates multiple estimations of the same modes. Theoretically, the model order should be set to K in each channel. However, this is usually a large over-estimation. Because each channel of the GT behaves like a band-pass filter, an EDS with a frequency far from $\frac{m}{M}$ will be attenuated and

considered as noise. Thus practically, the exact number of detectable components in each channel is unknown. The model order in each channel is therefore determined using the ESTER criterion (see section 3.4 for implementation details).

3.4. Discarding multiple components

If the distance between a set of channels on which an analysis has been performed is smaller than the bandwidth of the analysis window $g[l]$, the same components are likely to appear in all of these channels. These multiple estimations of the same component (hereafter named replicas) have to be identified. The only one that will be kept for the final re-synthesis is the one which frequency is the closest to the central frequency of the channel where it has been detected. A component c_r (with frequency f_r) is considered a replica of a component c_o (with frequency f_o) if the following conditions are fulfilled:

$$|f_r - f_o| < \varepsilon_f \quad (17)$$

$$|f_r - f_o| < |f_o - f_i| \quad (18)$$

Here ε_f is a frequency confidence interval and f_i is the closest frequency to f_r among the components detected in the same channel as c_r.

3.5. Discarding irrelevant components

Practical tests have shown that some of the modes detected using the previously describes approach are not relevant for they have an insignificant energy. An example of such a situation is covered in section 4. In order to produce satisfactory re-synthesis, it is important to take psycho-acoustical considerations into account. It is known that the human auditory system can be modelled as a band-pass filter bank. The filters bandwidths, called *critical bands*, are functions of the central frequency. In order not to favour any frequency range over an other in the discarding process, the idea is to keep only the components that added one to the other form 99% of the total energy in each of these filters. Therefore the following process is applied: first, the frequency domain is segmented in critical bands centred around each component. The energy of each critical band is then computed considering all the components which frequency falls into the given critical band. Finally, these components are added one to another, until 99% of the critical band energy is reached, and the remaining components are definitely discarded of the final re-synthesis. At every step of this process, the most energetic components are considered in priority.

4. NUMERICAL TESTS

4.1. Robustness to noise

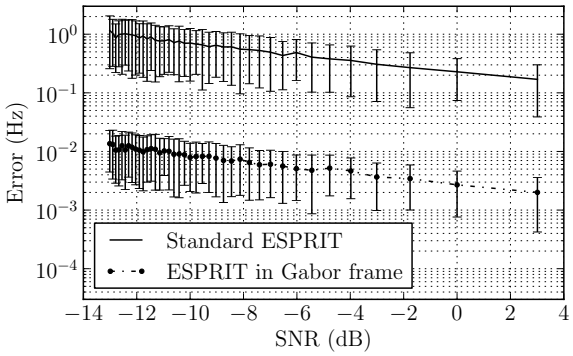

Fig. 1: Comparison of the frequency errors as a function of the SNR, with the ESPRIT method applied in the full-band time domain and with the ESPRIT method applied in a Gabor frame. For each SNR, 150 realisations have been computed.

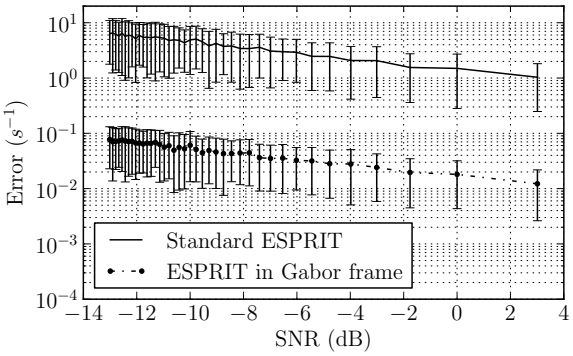

Fig. 2: Damping errors as a function of the SNR

This section compares the performances of the standard ESPRIT method against the proposed method, as the

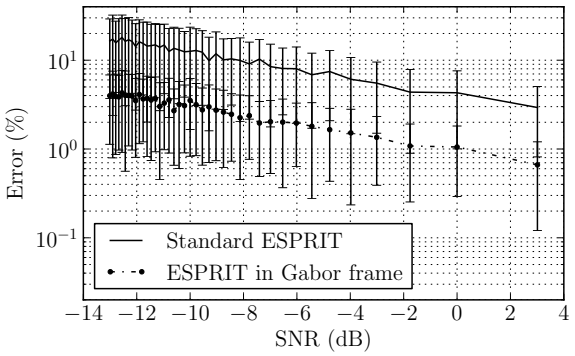

Fig. 3: Initial amplitude errors as a function of the SNR.

signal-to-noise-ratio (SNR) increases. The test signal is a sinusoid of frequency $f = 5000\ Hz$, damping $\delta = 1$, initial amplitude $A = 1$ and length $L = 70560$ (1.6s at 44100 Hz). The noise in which the test signal is embedded is a white noise of variance σ^2. For each value of σ^2, the SNR is computed at time $n = 0$ by:

$$20\log_{10}\left(\frac{A}{\sqrt{2}\sigma}\right) \tag{19}$$

At each SNR, 150 different realisations are computed. The Gabor frame consists in a blackman-harris window of length 2048, a time-parameter $a = 32$ and a number of channels $M = 2048$. For the full-band standard ESPRIT method, 2205 samples are analysed. For the Gabor frame method, the totality of the signal is analysed. This ensures that the resulting computational cost is of the same order in both cases ($2205 \times a = L$). Fig. 1 shows the error committed in the frequency estimation of the test signal. It shows that the frequency estimation error is significantly smaller in the Gabor frame case. Fig. 2 shows the estimated dampings for both methods. One can observe that the standard deviation to the real damping value is clearly higher in the full-band case. Similar observations can be made on the errors committed in the initial amplitude estimations Fig. 3. These tests clearly highlight the advantages of the Gabor frame approach; it is reasonable to suppose that the better behaviour observed in the Gabor frame case is due to the time sub-sampling, which increases the time-equivalent length of the analysed signal.

4.2. Robustness to damping

In order to test the robustness of both methods when the damping of the component increases, the following test has been elaborated. The deterministic part of the signal is an exponentially damped sinusoid of frequency 5000 Hz which damping varies from 1 to 100. The stochastic part is made of a white noise of variance 0.05, such that the SNR at the origin is 10 dB. This tests show that the Gabor frame approach is only superior to the full-band approach until a given damping threshold. The threshold value is around 23 for the estimation of the damping and the frequency (Fig. 4 and Fig. 5) and 15 for the initial amplitude estimation (Fig. 6). This corresponds to components which spend 99% of their energy in respectively 0.1 and 0.15 s. One can assume that for high dampings, the number of time-frequency transform coefficients which contains significant deterministic energy is too small for a correct estimation.

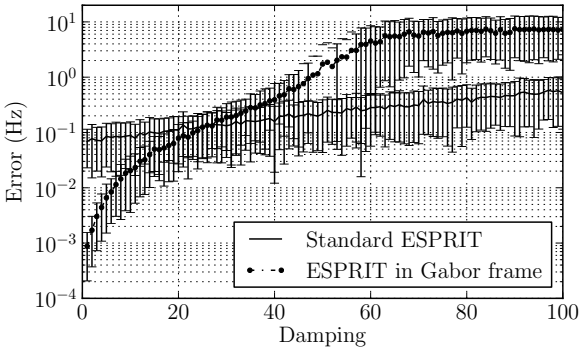

Fig. 4: Frequency errors as a function of damping.

5. APPLICATION ON A REAL-LIFE SOUND

This section focuses on the analysis/synthesis of a real sound `s1` (which can be listened to at [15]). `s1` corresponds to the sound of a metal bowl. Observing its spectrogram Fig. 6, one can see that it presents a rich spectral content and significant lasting energy up to 12 s.

5.0.1. Analysis with full-band ESPRIT method

Considering the size of the Hankel matrix corresponding the whole sound (around 260000×260000), only a part of the original signal can be analysed with the

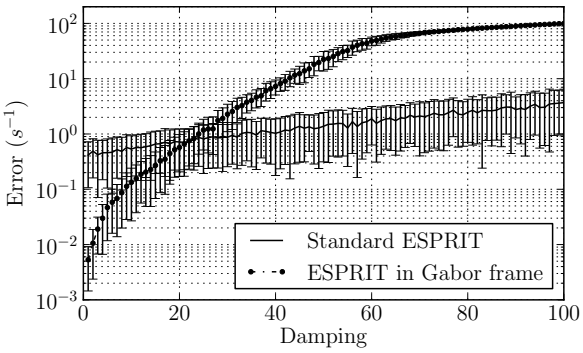

Fig. 5: Damping errors as a function of the damping.

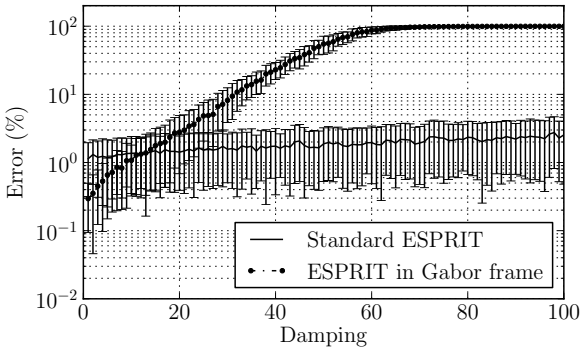

Fig. 6: Initial amplitude errors as a function of the damping.

full-band ESPRIT algorithm. Here, only the 10000 first samples of the initial signal are considered for the analysis. The order of the analysis is roughly overestimated at 300. After applying the ESPRIT algorithm, 8 EDS appear to have a negative damping, which will form diverging components when re-synthesising. Two different re-synthesis are proposed: the first obtained by arbitrarily setting the negative dampings to 1 (`s1_std_esprit_am_1.wav`), and the other one by discarding the components with a negative damping (`s1_std_esprit_am_sup.wav`). Their respective spectrograms are presented Fig. 8. The resulting synthesised sounds ([15]) are both unsatisfactory from a perceptual point of view. Fig. 8 shows that although some peak frequencies where correctly estimated by the standard ESPRIT method, the damping behaviour of the partials does not correspond to the one observed in the original sound (Fig. 6), especially for the components below 4000 Hz.

5.0.2. Analysis with ESPRIT in a Gabor transform

The chosen Gabor frame consists in a Blackman-Harris window of length 2048, a time-step parameter $a = 64$, and a number of channels $M = 2048$. It is unnecessary to apply the ESPRIT algorithm over regions of the time-frequency plane that only contain noise. Since the most important deterministic information is contained in the channels of high energy, these channels can be identified using a peak detection algorithm over the energy profile of the Gabor transform as shown in Fig. 9. In a software environment, the choice of which channels will be analysed could be left to the user. It is reasonable to think that the noise whitening induced by the subband division of the spectrum makes the ESTER criteria more reliable than in the full-band case, therefore the analysis order is computed for each of the selected channels, and set to the maximum of the ESTER criteria cost function. Doing so and after discarding the components which have a negative damping, a total number of 650 modes is obtained. Rejecting the irrelevant components as described in section 3.5, the number of components used for the final re-synthesis drops to 144. The damping of the components before and after applying the discarding process are shown Fig. 10, their amplitudes Fig. 11. The resulting re-synthesis `s1_esprit_gabor_650.wav` and `s1_esprit_gabor_144.wav` can be listened to at ([15]). The spectrogram of the final re-synthesis `s1_esprit_gabor_144.wav` is plotted Fig. 12. It can be noted that although peaks are missing in the final re-synthesis, it is much more satisfactory from a perceptual point of view than in the full-band case. It can be also observed that the global damping behaviour of the partials is much more similar to the original sound (Fig. 6).

6. FURTHER IMPROVEMENTS

One of the advantages provided by the use of time-frequency representations is the existence of efficient statistical estimators for the background noise. As it can be seen on Fig. 6, a significant number of Gabor coefficients describing an impact sound correspond to noise, and can therefore be used to estimate the variance of the stochastic part of the signal (see [11]). If the additive noise is

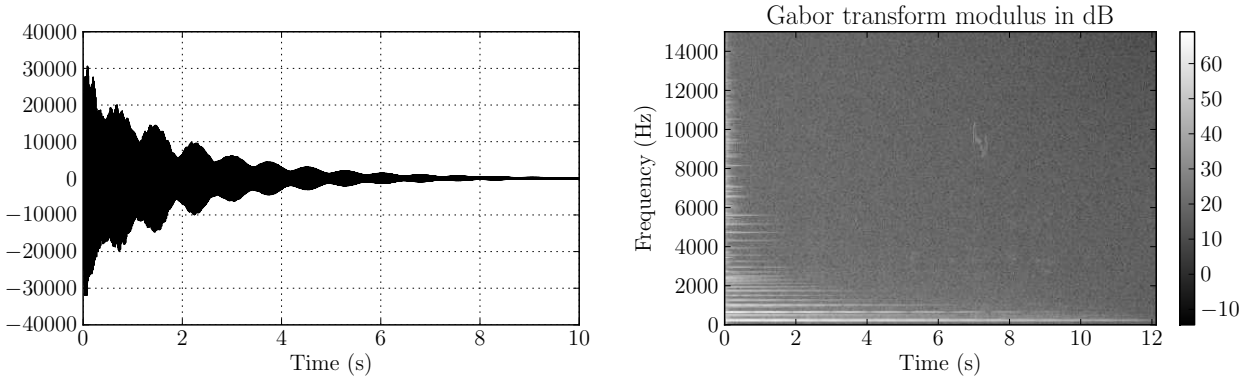

Fig. 7: Waveform of `s1.wav` on the left, and the corresponding spectrogram in dB on the right.

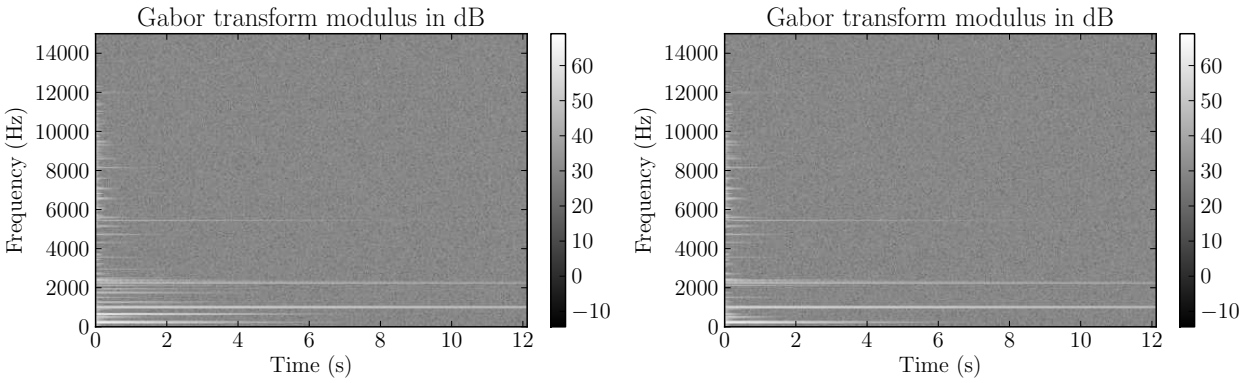

Fig. 8: Spectrograms of the the re-synthesised sounds after a standard ESPRIT analysis. The left figure corresponds to a re-synthesis after arbitrarily setting the negative dampings to 1 (`s1_std_esprit_am_1.wav`), and the right figure to a re-synthesis where the components presenting a negative damping have simply been discarded (`s1_std_esprit_am_sup.wav`).

coloured, it is even possible to estimate the variance in several selected frequency bands. Knowing the variance of the noise for each frequency channel offers the possibility to use noise masking properties of the human hearing to discard inaudible components, and possibly lead to a more selective criteria than the rejecting process described in section 3.5.

The concept of nonstationary Gabor frames ([13]) makes it also possible to adapt the resolution of the Gabor transform so as to get an optimal compromise between precision and computational cost. It would allow, for instance, to take into account the logarithmical frequency resolution of the human hearing when applying the Gabor transform. Furthermore, it can be observed that the damping usually decreases with frequency; nonstationary Gabor frames would allow to adapt the time-step parameter of the Gabor frame along the frequency scale, so that computational cost is saved while a sufficient number of coefficients are taken for the analysis.

7. CONCLUSION

It has been shown that using ESPRIT in time-frequency representations allows a better estimation of poles and amplitudes, except for very high dampings. The better robustness to noise of the Gabor frame approach has also been clearly highlighted. The consistency of the pro-

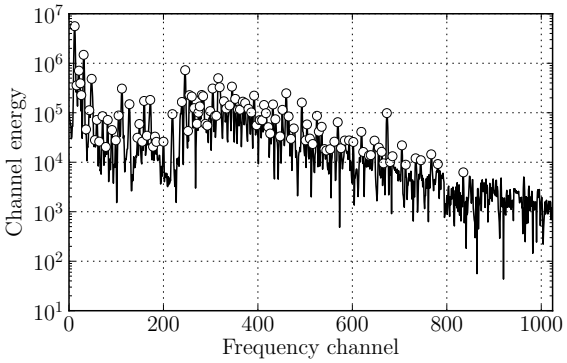

Fig. 9: Energy of the Gabor transform of `s1.wav` computed for each of its channels. The dots correspond to the 112 channels identified as peaks.

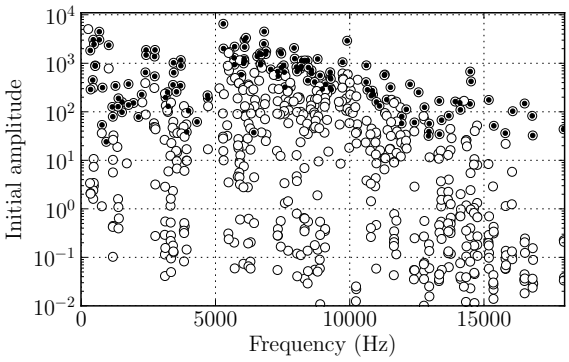

Fig. 11: Initial amplitudes of the components obtained after analysing `s1.wav`. The dotted components are the one which are kept after applying the discarding process described in section 3.5.

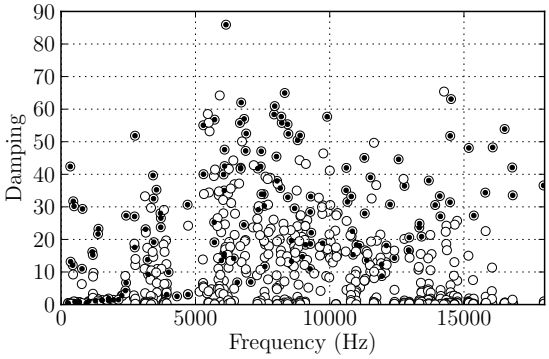

Fig. 10: Damping of the components obtained after analysing `s1.wav`. The dotted components are the one which are kept after applying the discarding process described in section 3.5.

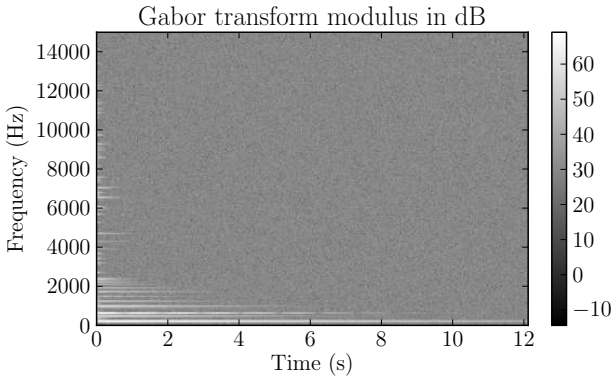

Fig. 12: Spectrogram of the re-synthesised sound `s1_esprit_gabor_144` (in black) resulting from an ESPRIT analysis within a Gabor frame, after discarding the irrelevant components.

posed method has been illustrated by an convincing resynthesis of a metallic sound. The Gabor frame approach has the same benefits as the sub-band analysis: it allows an extension of the analysis horizon, and it diminishes the complexity of the problem by only considering successive regions in the frequency domain; but on top of that, the information given by the time-frequency representation is of great use for targeting the analysis on the time-frequency intervals that contain the desired information. This avoids unnecessary analysis and reduces the global computational cost.

8. AKNOWLEDGMENTS

The authors would like to thank Julien Perron for his role in the calibration of the method. This project has been partly supported by the French National Research Agency (ANR-10-CORD-010 “Métaphores sonores”, http://metason.cnrs-mrs.fr/).

9. REFERENCES

[1] C. Verron, M. Aramaki, R. Kronland-Martinet and G. Pallone, "A 3-D immersive synthesizer for environmental sounds," Audio, Speech, and Language Processing, IEEE Transactions on, vol. 18, no. 6, pp. 1550-1561, 2010.

[2] "SoundDesignToolkit," http://www.soundobject.org/SDT/.

[3] R. Schmidt, "Multiple emitter location and signal parameter estimation," Antennas and Propagation, IEEE Transactions on, vol. 34, no. 3, pp. 276-280, 1986.

[4] R. Roy and T. Kailath, "ESPRIT - Estimation of Signal Parameters via Rotational Invariance Techniques," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 37, no. 7, pp. 984-995, 1989.

[5] M. Goodwin, "Matching Pursuits with Damped Sinusoids," Acoustics, Speech, and Signal Processing, 1997. Proceedings. (ICASSP'97). IEEE International Conference on, pp. 2037-2040, 2004.

[6] R. Badeau, "Méthodes haute-résolution pour l'estimation et le suivi de sinusoïdes modulées," Ph.D. Thesis, École Nationale Supérieure des Télécommunications, 2005

[7] K. Ege, X. Boutillon and B. David, "High-resolution modal analysis,"" Journal of Sound and Vibration, vol. 325, no. 4-5, pp. 852–869, 2009.

[8] J. Laroche, "A new analysis/synthesis system of musical signals using Prony's method-application to heavily damped percussive sounds," Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on, pp. 2053-2056, 1989.

[9] R. Badeau, B. David and G. Richard, A new perturbation analysis for signal enumeration in rotational invariance techniques, Signal Processing, IEEE Transactions on, vol. 54, no. 2, pp. 450–458, 2006.

[10] S. Van Huffel, H. Park and J.B. Rosen, "Formulation and solution of structured total least norm problems for parameter estimation," Signal Processing, IEEE Transactions on, vol. 44, no. 10, pp. 2464-2474, issn 1053-587X, 1996.

[11] F. Millioz and N. Martin, "Estimation of a white Gaussian noise in the Short Time Fourier Transform based on the spectral kurtosis of the minimal statistics: Application to underwater noise," Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, pp. 5638-5641, issn 1520-6149, 2010.

[12] T. Painter, A. Spanias, "Perceptual coding of digital audio," Proceedings of the IEEE, vol. 88, no. 4, pp. 451-515, issn 0018-9219, 2000.

[13] F. Jaillet, P. Balazs and M. Dörfler, "Nonstationary Gabor frames," Proceedings of the 8th international conference on Sampling Theory and Applications (SAMPTA'09), May 2009.

[14] K. Gröchenig, "Foundations of time-frequency analysis," Birkhauser, 2001.

[15] Sample sounds, available at: http://www.lma.cnrs-mrs.fr/~kronland/AES2012/.