

## **EDS parametric modeling and tracking of audio signals**

Roland Badeau, Rémy Boyer, Bertrand David

► **To cite this version:**

Roland Badeau, Rémy Boyer, Bertrand David. EDS parametric modeling and tracking of audio signals. Proc. of the 5th International Conference on Digital Audio Effects (DAFx), 2002, Hambourg, Germany. pp.139–144, 2002. <hal-00945272>

**HAL Id: hal-00945272**

**<https://hal.inria.fr/hal-00945272>**

Submitted on 24 Mar 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## EDS PARAMETRIC MODELING AND TRACKING OF AUDIO SIGNALS

Roland Badeau, Rémy Boyer and Bertrand David

ENST, Département de Traitement du Signal et des Images  
46, rue Barrault, 75634 Paris Cedex 13 France  
roland.badeau, remy.boyer, bertrand.david@enst.fr

### ABSTRACT

Despite the success of parametric modeling in various fields of digital signal processing, the Fourier analysis remains a prominent tool for many audio applications. This paper aims at demonstrating the usefulness of the Exponentially Damped Sinusoidal (EDS) model both for analysis/synthesis and tracking purposes.

### 1. INTRODUCTION

The advantages of this model are on the one hand to overcome the Fourier resolution limit related to windowing and on the other hand to enhance the classical sinusoidal model used in speech and audio coding. The main drawbacks of the EDS-based methods consist in the complexity of the algorithms and in the assumption of non time-varying parameters. In this paper, applications of some recent enhancements of the algorithms to audio signal processing are presented, in order to both reduce the complexity and track the parameter variations with time.

**Modeling context.** Recently, many efforts have been made to achieve a powerful representation of an audio signal as speech or music, for a compression purpose [1, 2]. More specifically, in parametric audio coding, it is worthwhile to have compact (sparse) representations of the signal: the model order  $M$  (*i.e.* the number of elementary components) must be far less than  $N$ , the length of the analysis window in samples. One way to obtain a more compact representation is to increase the parameter  $N$  and keep unchanged the model order. Unfortunately, for large  $N$ , the audio signal can no longer be considered as a quasi-stationary signal. In this case, the basic sinusoidal model [3], which tends to represent the audio signal as a sum of constant-amplitude components, becomes ineffective. Consequently, the EDS model was introduced in the audio modeling context [4, 5]. In this work, we compare the sinusoidal and EDS models with the same total number of model parameters. Note that keeping a satisfactory algorithmic complexity implies setting a maximal bound to the parameter  $N$ .

**Tracking context.** The EDS model relies on the assumption of non varying signal parameters within the observation window. A more realistic modeling of musical signals should include slow variations of the parameters. Tracking these time variations would have interesting applications, such as:

- evaluating the degree of stationarity of the audio signal,
- detecting model breaks, which characterize transient sounds,
- developing more realistic synthesis techniques.

A reference method in frequency tracking is the Sintrack algorithm introduced by P. Duvaut [6]. This method relies on a fast linear prediction technique, which makes it useful for real-time estimation and tracking of damped sinusoids in noise. However, its lack

of robustness results in repeated re-initializations which increase the computational cost.

Concurrently, subspace-based high resolution methods, despite their higher computational complexity, prove to be much more reliable than linear prediction. Therefore, adaptive subspace estimation may offer interesting outlooks for frequency tracking.

**Contents.** This paper is organized as follows.

Section 2 introduces the EDS model and presents subspace-based high resolution methods for the estimation and tracking of the model parameters. Some synthesis techniques are proposed both in a static and an adaptive context, with an application to pitch modification.

Section 3 shows the application of these methods to coding, tracking and re-synthesis of audio signals.

Finally, section 4 summarizes the main conclusions of this paper.

### 2. THEORETICAL BACKGROUND

The EDS model defines the discrete signal as

$$x(t) = \sum_{m=1}^M a_m \exp(d_m t) \cos(2\pi f_m t + \phi_m), \quad t \in \{0, \dots, N-1\} \quad (1)$$

where  $x(t)$  is the discrete signal observed in the window  $t \in \{0, \dots, N-1\}$ ,  $M$  is the order of the model,  $a_m \in \mathbb{R}_+$  are the amplitudes,  $d_m$  are the real valued damping factors,  $f_m \in [-\frac{1}{2}, \frac{1}{2}[$  are the frequencies and  $\phi_m \in [-\pi, \pi[$  denote initial phases. Equation (1) can equivalently be rewritten with the complex amplitudes  $\alpha_m = \frac{1}{2} a_m \exp(i\phi_m)$  and the complex poles  $z_m = \exp(d_m + i2\pi f_m)$  as in equation (2):

$$x(t) = \sum_{m=1}^M (\alpha_m z_m^t + \alpha_m^* z_m^{*t}). \quad (2)$$

In section 2.1, EDS-based analysis/synthesis methods are presented in a block processing context (with constant model parameters). In section 2.2, it will be shown how these methods can be adapted to track slow variations of these parameters.

#### 2.1. Block signal processing

The estimation of the model parameters is achieved in two steps: first the frequencies and damping factors are computed using a high resolution (HR) method, from which the amplitudes and initial phases are deduced by minimizing a least squares (LS) criterion. The estimated parameters are then used to re-synthesize the signal.

Table 1: Orthogonal iteration EVD algorithm

---

Initialization:  $U^S = \begin{bmatrix} \mathbf{I}_{2M} \\ \mathbf{0}_{(L-2M) \times 2M} \end{bmatrix}$   
 For  $n = 1, 2, \dots$  until convergence iterate:

---

$\mathbf{A}(n)$	$=$	$\mathbf{H} \mathbf{U}^S(n-1)$	fast matrix product
$\mathbf{A}(n)$	$=$	$\mathbf{U}^S(n) \mathbf{R}(n)$	skinny QR factorization

---

### 2.1.1. Subspace-based signal analysis

Define the  $L \times L$  real Hankel data matrix  $\mathbf{H}$  (with  $N = 2L - 1$ ) as

$$\mathbf{H} = \begin{bmatrix} x(0) & x(1) & \dots & x(L-1) \\ x(1) & x(2) & \dots & x(L) \\ \vdots & \vdots & \ddots & \vdots \\ x(L-1) & x(L) & \dots & x(N-1) \end{bmatrix}. \quad (3)$$

Suppose that  $2M \leq L$ . Then this matrix can be decomposed as  $\mathbf{H} = \mathbf{E} \mathbf{A} \mathbf{E}^T$ , where

$$\mathbf{A} = \text{Diag}(\alpha_1, \dots, \alpha_M, \alpha_1^*, \dots, \alpha_M^*)$$

and  $\mathbf{E}$  is the  $L \times 2M$  Vandermonde matrix

$$\mathbf{E} = \begin{bmatrix} 1 & \dots & 1 & 1 & \dots & 1 \\ z_1 & \dots & z_M & z_1^* & \dots & z_M^* \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_1^{L-1} & \dots & z_M^{L-1} & z_1^{*L-1} & \dots & z_M^{*L-1} \end{bmatrix}. \quad (4)$$

$\mathbf{H}$  has a  $2M$ -dimensional range space, spanned by the full-rank matrix  $\mathbf{E}$ . This range space fully characterizes the signal poles, even in presence of an additive white noise [7], and thus is referred to as the *signal subspace*. An orthonormal basis  $\mathbf{U}^S$  of this space can be obtained from the eigenvalue decomposition (EVD) of  $\mathbf{H}$ . Indeed, since  $\mathbf{H}$  is a rank-deficient symmetric real matrix, there exist a  $L \times 2M$  orthonormal real matrix  $\mathbf{U}^S$  and a  $2M \times 2M$  diagonal real matrix  $\mathbf{\Lambda}$  such that  $\mathbf{H} = \mathbf{U}^S \mathbf{\Lambda} \mathbf{U}^{S^T}$ . The columns of  $\mathbf{U}^S$  thus span the signal subspace. In the presence of an additive white noise, the columns of  $\mathbf{U}^S$  are defined as the  $2M$ -dominant eigenvectors of  $\mathbf{H}$  (i.e. the eigenvectors associated to the  $2M$  eigenvalues which have the highest magnitudes).

These dominant eigenvectors can be computed using the classical EVD algorithm called *orthogonal iteration* [8]<sup>1</sup> (cf. table 1), which involves an auxiliary matrix  $\mathbf{A}$ . The Hankel structure of the matrix  $\mathbf{H}$  can be taken into account to make the algorithm faster by computing the first-step matrix product using Fast Fourier Transforms, which requires only  $O(LM \log(L))$  operations [8]<sup>2</sup>. Then the second step can be achieved in  $O(LM^2)$  operations [8]<sup>3</sup>. Since in practice this algorithm converges in a few iterations, the overall process requires  $O(LM(M + \log(L)))$  operations.

<sup>1</sup>Chapter 8, section 2.4.

<sup>2</sup>Chapter 4, section 7.7.

<sup>3</sup>Chapter 5, section 2.

### 2.1.2. Estimation of the frequencies and damping factors

The poles  $\{z_m, z_m^*\}_{1 \leq m \leq M}$  can be calculated by exploiting the rotational invariance property of the signal subspace. More precisely, define  $\mathbf{E}_\downarrow$  (respectively  $\mathbf{E}_\uparrow$ ) the matrix extracted from  $\mathbf{E}$  by deleting the last (respectively the first) row. These matrices satisfy the equation

$$\mathbf{E}_\uparrow = \mathbf{E}_\downarrow \mathbf{D} \quad (5)$$

where

$$\mathbf{D} = \text{diag}(z_1, z_2, \dots, z_M, z_1^*, z_2^*, \dots, z_M^*).$$

Since the matrices  $\mathbf{U}^S$  and  $\mathbf{E}$  span the same subspace, there exist an invertible matrix  $\mathbf{C}$  such that

$$\mathbf{U}^S = \mathbf{E} \mathbf{C}^{-1}. \quad (6)$$

As for  $\mathbf{E}$ , let  $\mathbf{U}_\downarrow^S$  (respectively  $\mathbf{U}_\uparrow^S$ ) be the matrix extracted from  $\mathbf{U}^S$  by deleting the last (respectively the first) row. Then equations (5) and (6) yield

$$\mathbf{U}_\uparrow^S = \mathbf{U}_\downarrow^S \mathbf{\Phi} \quad (7)$$

where  $\mathbf{\Phi} = \mathbf{C} \mathbf{D} \mathbf{C}^{-1}$ . The Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) method [9] consists in:

- computing the matrix  $\mathbf{\Phi} = (\mathbf{U}_\downarrow^S)^\dagger \mathbf{U}_\uparrow^S$  (where the symbol  $\dagger$  denotes the Moore-Penrose pseudo-inverse; this computation requires  $O(LM^2)$  operations),
- extracting the estimated poles  $\hat{z}_m$  as the eigenvalues of  $\mathbf{\Phi}$  (which can be achieved in  $O(M^3)$  operations).

Finally, for  $m = 1, \dots, M$ , the  $m^{\text{th}}$  estimated frequency and damping factor can be deduced using  $\hat{f}_m = \frac{\text{angle}(\hat{z}_m)}{2\pi}$  and  $\hat{d}_m = \ln |\hat{z}_m|$ .

### 2.1.3. Estimation of the amplitudes and initial phases

The complex amplitudes  $\{\alpha_m\}_{1 \leq m \leq M}$  can be determined by minimizing the LS criterion  $\min_{\alpha} \|\mathbf{x} - \mathbf{E} \alpha\|_2^2$ , where

- $\mathbf{x} = [x(0), \dots, x(L-1)]^T$  are the signal samples,
- $\alpha = [\alpha_1, \dots, \alpha_M, \alpha_1^*, \dots, \alpha_M^*]^T$  are complex amplitudes.

The solution to this criterion is

$$\hat{\alpha} = \mathbf{E}^\dagger \mathbf{x}. \quad (8)$$

Hence, for  $m = 1, \dots, M$ , the  $m^{\text{th}}$  estimated real amplitude and initial phase are  $\hat{a}_m = 2|\hat{\alpha}_m|$  and  $\hat{\phi}_m = \text{angle}(\hat{\alpha}_m)$ .

Note that the full computation of  $\mathbf{E}^\dagger$  can be avoided since equation (6) shows that  $\mathbf{E}^\dagger = \mathbf{C}^{-1} \mathbf{U}^{S^T}$  where  $\mathbf{C} = \mathbf{U}^{S^T} \mathbf{E}$ . Thus,  $\hat{\alpha} = \mathbf{C}^{-1} (\mathbf{U}^{S^T} \mathbf{x})$  can be computed in  $O(LM^2)$  operations.

### 2.1.4. Re-synthesis

Once the model parameters have been estimated, the signal can be reconstructed using equation (2). Thus, the estimated signal sample at time  $t$  is

$$\hat{x}(t) = \sum_{m=1}^M (\hat{x}_m(t) + \hat{x}_m^*(t)) \quad (9)$$

where  $\hat{x}_m(t) = \hat{\alpha}_m \hat{z}_m^t$  is the  $m^{\text{th}}$  complex damped sinusoid.

Note that equation (9) can be implemented in  $O(LM)$  operations.

In a block processing context, some interpolation techniques are required in order to force the continuity of the parameters between consecutive blocks [10].

### 2.1.5. Pitch-scale modification

An immediate application of the EDS model is a frequency-scale modification of the signal, which just consists in multiplying the estimated frequencies  $\hat{f}_m$  by a same factor  $\beta$ . Thus, the frequency of the  $m^{\text{th}}$  complex damped sinusoid in the modified signal is

$$\hat{f}_m^s = \beta \hat{f}_m$$

so that the corresponding pole is

$$\begin{aligned} \hat{z}_m^s &= \exp(\hat{d}_m + i 2\pi \hat{f}_m^s) \\ &= \hat{z}_m \exp(i 2\pi (\beta - 1) \hat{f}_m) \end{aligned} \quad (10)$$

Therefore, equation (9) becomes

$$\hat{x}^s(t) = \sum_{m=1}^M (\hat{x}_m^s(t) + \hat{x}_m^{s*}(t)) \quad (11)$$

where  $\hat{x}_m^s(t) = \hat{\alpha}_m (\hat{z}_m^s)^t$ .

Note that this pitch modification method is no more computationally demanding than the exact re-synthesis.

## 2.2. Adaptive signal processing

The section transposes the HR methods presented above in an adaptive context. It will be shown that tracking the slow variations of the model parameters leads to a very simple re-synthesis method.

### 2.2.1. Model parameters tracking

The Sintrack method for frequency estimation and tracking [6] consists in a two-steps estimation: the Matrix Pencil HR method [7] is first applied to obtain the initial parameters, and the tracking is then achieved using an adaptive Least Mean Square (LMS) algorithm, the frequencies and damping factors being extracted from the roots of a backward prediction polynomial [11]. When the prediction error exceeds a certain threshold, the algorithm switches back to the initialization step.

Although this method has proved to be successful on musical signals [12], the lack of robustness of the LMS algorithm results in an intensive use of the Matrix Pencil method, which is very time-consuming.

To avoid this increase of complexity, the prediction polynomial tracking can be replaced by a signal subspace tracking, since subspace-based HR methods are known to give more reliable estimates of the signal poles than linear prediction. Subspace tracking has been intensively studied in the fields of adaptive filtering, source localization or parameter estimation. A first class of tracking algorithms is based on the projection approximation hypothesis [13]; an other one relies on EVD or SVD tracking techniques, derived from classical EVD or SVD algorithms.

Table 2: Sequential iteration EVD algorithm

Initialization:	$U^S = \begin{bmatrix} \mathbf{I}_{2M} \\ \mathbf{0}_{(L-2M) \times 2M} \end{bmatrix}$
For each time step $t$ iterate:	
$\begin{bmatrix} \mathbf{A}(t) \\ \mathbf{A}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{H}(t) \mathbf{U}^S(t-1) \\ \mathbf{U}^S(t) \mathbf{R}(t) \end{bmatrix}$	fast matrix product skinny QR factorization

For example, the *orthogonal iteration* algorithm of Table 1 can be adapted to track the dominant eigenvectors of a sliding-window matrix

$$\mathbf{H}(t) = \begin{bmatrix} x(t-(L-1)) & \dots & x(t) \\ x(t-(L-2)) & \dots & x(t+1) \\ \vdots & \ddots & \vdots \\ x(t) & \dots & x(t+L-1) \end{bmatrix} \quad (12)$$

just by replacing the iteration index  $n$  in table 1 by the discrete time index  $t$  [14] (cf. table 2).

Thus, only one iteration is completed at each time step. Once the signal subspace basis  $U^S$  is computed, the standard ESPRIT method can be applied. However, for the sake of computational efficiency, adaptive implementations of ESPRIT have been developed [15], which require  $O(LM^2)$  or  $O(LM)$  operations at each time step.

Finally, the estimation of the amplitudes and initial phases can be achieved as in section 2.1.3. Equation (8) now becomes

$$\hat{\alpha}(t) = \mathbf{E}(t)^\dagger \mathbf{x}(t) \quad (13)$$

where  $\mathbf{E}(t)$  is the Vandermonde matrix of the estimated poles at time  $t$ ,  $\hat{\alpha}_m(t)$  and  $\hat{z}_m(t)$  denote the estimated  $m^{\text{th}}$  complex amplitude and pole at time  $t$ , and  $\mathbf{x}(t) = [x(t), \dots, x(t+L-1)]^T$ .

Since this estimation involves the matrix  $\mathbf{E}$  defined in equation (4) for a time window  $[0 \dots L-1]$ , it must be noted that  $\hat{\alpha}_m(t)$  now is the complex amplitude of the  $m^{\text{th}}$  damped sinusoid at time  $t$ .

### 2.2.2. Re-synthesis

In an adaptive context, since the complex amplitudes of the damped sinusoids are estimated at each time step, equation (9) stands with  $\hat{x}_m(t) = \hat{\alpha}_m(t)$ . Therefore, the re-synthesis of the signal at each time step just consists in summing the complex amplitudes, which only requires  $O(M)$  operations.

### 2.2.3. Pitch scale modification

Let  $\varphi_m(t)$  be the phase shift between the  $m^{\text{th}}$  estimated damped sinusoid and the  $m^{\text{th}}$  synthesized damped sinusoid at time  $t$ , so that equation (11) stands with

$$\hat{x}_m^s(t) = \hat{x}_m(t) \exp(i \varphi_m(t)) = \hat{\alpha}_m(t) \exp(i \varphi_m(t)). \quad (14)$$

Since these sinusoids satisfy the following recurrences

- $\hat{x}_m(t) = \hat{x}_m(t-1) \hat{z}_m(t)$ ,

- $\hat{x}_m^s(t) = \hat{x}_m^s(t-1) \hat{z}_m^s(t)$ ,

equations (10) and (14) show that  $\varphi_m(t)$  can be recursively updated using the following scheme:

$$\varphi_m(t) = \varphi_m(t-1) + 2\pi(\beta-1)\hat{f}_m(t). \quad (15)$$

Then,  $\hat{x}_m^s(t)$  can be computed using equation (14), from which the synthesized sample  $\hat{x}^s(t)$  can be deduced using equation (11). Note that this pitch modification method has the same complexity as the exact re-synthesis.

### 3. EXPERIMENTAL RESULTS

This section illustrates first the enhancement of the coding quality using the EDS rather than a simple sinusoidal model, then the tracking and re-synthesis of musical signals. The study deals with two piano tones, C5 and G5, sampled at 11025 Hz.

It is to be noticed that in real audio signal applications, the data matrix  $\mathbf{H}$  is never rank-deficient, because of the presence of noise. Moreover, the rank-truncation order  $2M$  is unknown, and must be chosen carefully. Indeed, over-estimating  $M$  is harmless, but under-estimating  $M$  often generates biases in the estimates of the frequencies and damping factors. Then,  $L$  must be chosen much greater than  $M$ , in order to enforce the robustness of the HR method. On the other hand, the higher  $L$  is, the more this method is computationally demanding. Therefore, audio signals with a high number of sinusoids (typically low-pitched sounds) may first be decomposed into several sub-band signals (via filtering/decimating, as proposed in [16]), before applying the HR method. In the examples proposed below, this pre-processing is useless, since the chosen piano tones have few sinusoidal components.

#### 3.1. EDS vs sinusoids

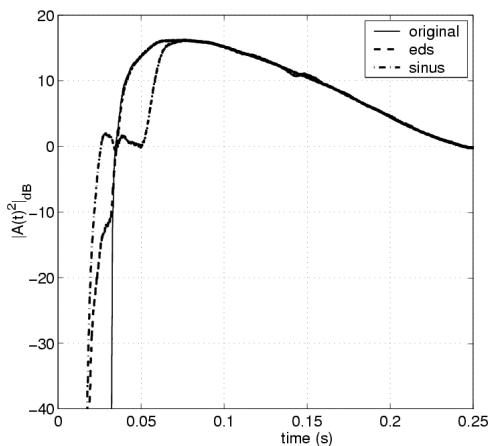


Figure 1: Time-shape waveforms : original signal, sinusoidal modeling with  $M_{SIN} = 16$  and EDS modeling with  $M_{EDS} = 12$ .

This section shows the efficiency of the EDS model in comparison with the classical sinusoidal model with an identical total number of model parameters, *i.e.*  $M_{EDS} = 3/4 M_{SIN}$ . The test signal is the C5 piano tone. Figures 1 and 2-a,b,c show the time-shape waveforms and the Fourier spectra of the original and

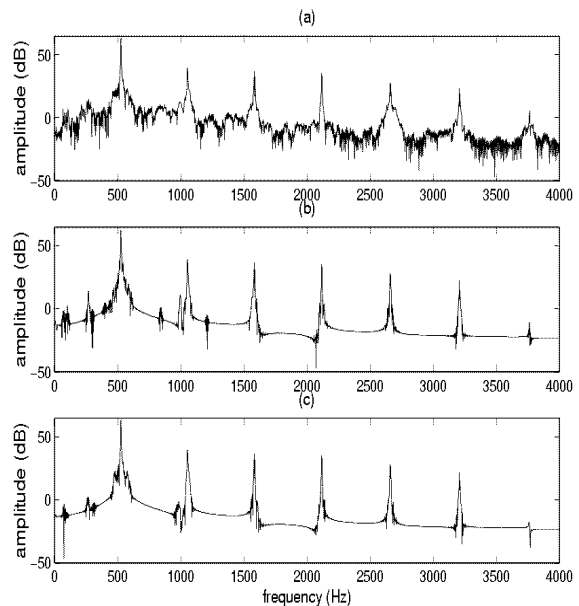


Figure 2: Fourier spectra : (a) original signal, (b) sinusoidal modeling with  $M_{SIN} = 16$ , (c) EDS modeling with  $M_{EDS} = 12$ .

modeled signals. Figure 1 shows a strong pre-echo (energy before the onset) with the sinusoidal model. Moreover, the global variation of the attack is wrongly estimated.

Thanks to the exponentially time-varying amplitudes, the EDS model provides a better modeling since it creates a short pre-echo and offers a good reproduction of the attack.

After several structural considerations, a frequency aspect is introduced in the analysis by using the polyphase 32-bands pseudo-QMF filter-bank of MPEG1-audio [17], which provides a uniform partition of the frequency axis. After that, a power and a SNR measure are computed in each sub-band, noted  $SNR_{TF}$ .

Figure 3 shows the better  $SNR_{TF}$  values of the EDS model. Note that several sub-bands are not reliable due to their weak power (see figure 4).

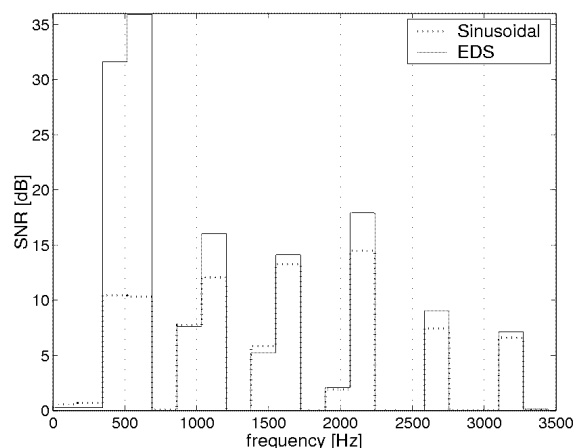


Figure 3:  $SNR_{TF}$  in dB.

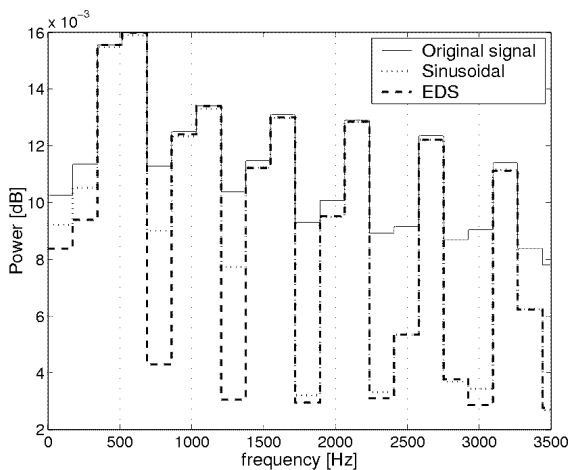


Figure 4: Power in sub-bands in dB.

### 3.2. Adaptive signal processing

This section illustrates first the frequency tracking of a musical signal composed of two piano tones, then the synthesis method of section 2.1.4.

#### 3.2.1. Frequency tracking

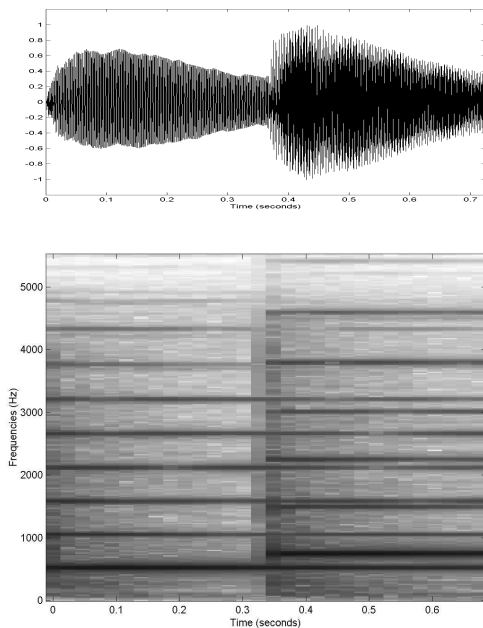


Figure 5: Time waveform and spectrogram of the piano signal for the frequency tracking test.

The parameters tracking method presented in section 2.2.1 has been tested on a piano signal: the C5 tone of figure 1-a is played at time  $t = 0s$ , then the G5 is played at time  $t = 0.36s$ , while the C5 is maintained (the time waveform and the spectrogram of

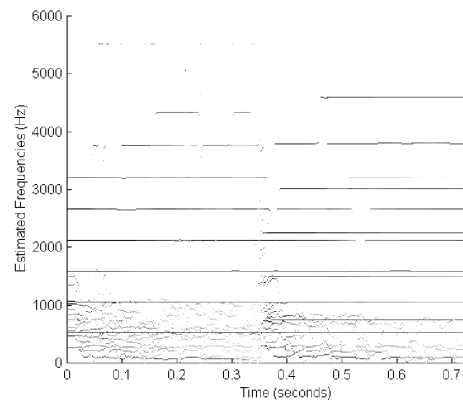


Figure 6: Frequency tracking of the piano signal.

this signal are plotted in figure 5). Figure 6 shows the result of the tracking. The model order and the window length where  $M = 16$  and  $L = 160$ , and the sinusoids energies ( $\mathcal{E}_m = \alpha_m^2 \frac{1-|z_m|^{2L}}{1-|z_m|^2}$ ) are represented on a logarithmic scale using gray levels for the plot. Since the number of sinusoids is over-estimated, it can be seen that spurious poles are detected in the low frequency band (below 1000Hz), which actually corresponds to the highest level of noise in the original signal.

#### 3.2.2. Re-synthesis and pitch scale modification

The synthesis method proposed in section 2.1.4 gave excellent results on the piano tones: the synthesized sounds were perceptually very similar to the original ones.

The hearing sensation is particularly well reproduced at the attack of both sounds. This may be related to the spurious poles detection mentioned above. Their number and energy are greater at the attack, which allows a good representation of the mechanical noise. It is well known that this impact noise occurring during the action of the hammer on the strings is of great importance for the naturalness of the sound.

This method could be directly implemented without any further modification. On the opposite, the pitch-scale modification requires additional work. Indeed, the recursion on the phase shift between the estimated and the pitch-shifted signal in equation (15) relies on several implicit assumptions, such as:

- the number of frequencies is constant through time,
- each pole characterizes one single time-varying frequency, which is present in the whole signal,
- $\hat{z}_m(t)$  matches  $\hat{z}_m(t-1)$  (ie. the  $m^{\text{th}}$  frequency trajectory is known).

In real audio signals, however, the frequencies may appear or disappear, so that their number changes through time. Moreover, spurious frequencies are sometimes detected, and should be eliminated. Consequently, tracking the poles trajectories is a difficult problem.

In the literature, several strategies were proposed to track sinusoids in the presence of noise in a block processing context [3, 10, 18]. These methods were designed in association with frequency estimators based on the Short Time Fourier Transform

(STFT), but they can easily be adapted to the EDS model and the HR methods.

Finally, the pitch-scale modification technique proposed in section 2.2.3 in combination with these classical frequency matching strategies proved to be successful on the piano tones.

Note that once the poles trajectories are estimated, the discrimination should be made between

- the harmonics (related to the pitch of the sound),
- the remaining poles, which model the signal noise.

A realistic pitch scale modification should change the frequencies of the first class and leave the second class unchanged. Of course, the classification of the poles would require additional work.

#### 4. CONCLUSIONS

The EDS model is a useful tool for audio signals modeling. It leads to a better representation of signal frames than the undamped sinusoidal model for a coding purpose. The use of a HR algorithm achieves an accurate estimation which can be efficiently updated by tracking the signal subspace through time. Moreover, tracking the model parameters offers very interesting outlooks for signal re-synthesis and modification.

#### 5. REFERENCES

- [1] ISO-MPEG, "Call for proposals for new tools for audio coding," *ISO/IEC JTC1/SC29/WG11 MPEG2001/N3793*, 2001.
- [2] H. Purnhagen and N. Meine, "HILN-the MPEG-4 parametric audio coding tools," in *Proc. of IEEE Int. Symposium on Circuits and Systems*, 2000.
- [3] R.J. McAulay and T.F. Quatieri, "Speech analysis and synthesis based on a sinusoidal representation," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. 34, no. 4, 1986.
- [4] J. Nieuwenhuijse, R. Heusdens, and E.F. Deprettere, "Robust exponential modeling of audio signal," in *Proc. of IEEE Int. Conf. on Acoustic, Speech and Signal Proc.*, May 1998.
- [5] R. Boyer, S. Essid, and N. Moreau, "Non-stationary signal parametric modeling techniques with an application to low bitrate audio coding," in *Proc. of IEEE Int. Conf. on Signal Proc.*, 2002.
- [6] Patrick Duvaut, *Traitement du signal*, Hermes, Paris, 1994.
- [7] Y. Hua and T.K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, May 1990.
- [8] G.H. Golub and C.F. Van Loan, *Matrix computations*, The Johns Hopkins University Press, Baltimore and London, third edition, 1996.
- [9] R. Roy and T. Kailath, "ESPRIT—estimation of signal parameters via rotational invariance techniques," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. 37, no. 7, 1989.
- [10] X. Serra and J. Smith, "Spectral modeling synthesis : a sound system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, 1990.
- [11] R. Kumaresan and D.W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 6, 1982.
- [12] B. David, R. Badeau, and G. Richard, "Sintrack analysis for tracking components of musical signals," in *Proc. of Forum Acusticum Sevilla 2002*, accepted for publication.
- [13] K. Abed-Meraim, A. Chkeif, and Y. Hua, "Fast orthonormal PAST algorithm," *IEEE Signal Processing Letters*, vol. 7, no. 3, 2000.
- [14] P. Strobach, "Square hankel SVD subspace tracking algorithms," *Signal Processing*, vol. 57, no. 1, 1997.
- [15] P. Strobach, "Fast recursive subspace adaptive ESPRIT algorithms," *IEEE Trans. on Signal Proc.*, vol. 46, no. 9, 1998.
- [16] J. Laroche, "The use of the Matrix Pencil method for the spectrum analysis of musical signals," *Journal of the Acoustical Society of America*, vol. 94, no. 4, 1993.
- [17] K. Banderburg and G. Stoll, "ISO-MPEG-1 audio: a generic standard for coding of high-quality digital audio," *Journal of the Acoustical Society of America*, vol. 42, 1994.
- [18] S. Levine, *Audio representations for data compression and compressed domain processing*, Ph.D. thesis, Stanford University, 1998.