Audio Modeling based on Delayed Sinusoids

Remy Boyer, Karim Abed-Meraim

Audio Modeling based on Delayed Sinusoids

Rémy Boyer and Karim Abed-Meraim

Abstract— **In this work, we present an evolution of the DDS (Damped & Delayed Sinusoidal) model introduced within the framework of the general signal modeling. This model is named the Partial Damped & Delayed Sinusoidal (PDDS) model and takes into account a single time delay parameter for a set (sum) of damped sinusoids. This modification is more consistent with the transient *audio* modeling problem. We show the validity of this approach by comparison with the well-known EDS (Exponentially Damped Sinusoids) approach. Finally, the performances of three model high-resolution parameter estimation algorithms are compared on synthetic fast time-varying signals and on two typical audio transients.**

Keywords— **Transient audio compact representations, damped and delayed sinusoids, high-resolution method**

I. Introduction

DURING the decade, many efforts have been made to achieve an efficient parametric representation of an audio signal for very low bite-rate compression purposes [1]. More precisely, the audio transient compact representation by parametric models is an up-to-date and difficult problem [2], [3], [4]. Basically, in an audio signal, we have two important features : the spectral content and the time waveform. For some kind of quasi-stationary signals, the most important is to well represent their spectral variation without considering too much the variation of the time waveform [3], [4], [5]. In the context of transient audio compact modeling, the signal time waveform is the main audio feature and have to be represented as best we can, *i.e.*, with minimum modeling errors. In the sequel, we define a transient signal as a signal whose support duration is short compared to the analysis range.

Parametric EDS (Exponentially Damped Sinusoidal) model has been widely studied in the signal processing community [6], [7], [8]. However, its application to signal compression is quite recent [9], [10], [12], [13], [14], [15], [16], [17]. This approach comes as a natural evolution of the sinusoidal model introduced by McAulay & Quatieri [5]. In fact, sinusoidal models assume that model parameters have slow variation regarding the analysis time range. Yet, this is not always consistent with the previous transient signal definition and when processing such diverse audio signal as speech or music.

EDS model and its extensions [10], [11] permit more appropriate fast time-varying signal modeling since each sinusoidal component amplitude is allowed to vary exponentially over time. Based on this property, these models present a growing interest in the audio community since they lead to compact (sparse) representations for almost the totality of audio signal. However, this model becomes ineffective on sharp transient signals like some percussive sounds (castanets, gong, triangle, ...) [9], [18], [19]. Modeling characteristic artifacts are created with two effects. First, the apparition of a pre-echo signal [4], [20], *i.e*, a distortion before the sound onset. Second, the signal dynamic is badly reproduced. These phenomena appear to be very prejudicial to the auditory perception of this sound category. Moreover, the onset part is of extreme importance for the "naturalness" of the audio signal [21].

Many approaches have been considered to solve this problem. These can basically be classified in four categories : the first is based on an irregular segmentation of the time axis [19], [22], the second exploits the time-frequency duality principle and the parametric modeling of a frequency-transformed signal [23], [24] and the third uses the "Matching-Pursuit" algorithm and the "Atomic" formalization to expand the signal on a redundant family (Gabor, EDS, ...) [15], [16]. Finally, an original method is presented in [13]. Recently, the parametric model, called DDS (Damped & Delayed Sinusoids) was presented in [18] as a generalization of the sinusoidal and EDS models. In this work, we make two realistic assumptions :

(A.1) A percussive audio signal can be seen as a set (sum) of damped sinusoids, all having a same time-delay.

(A.2) Two successive audio transients are at "sufficient" relative distance one from an other to perform an efficient time-delay estimation/detection based on the signal envelop variation.

In this context, we modify the general DDS model and introduce the Partial Damped & Delayed Sinusoidal (PDDS) model. This model can be seen as a generalization of the EDS model and a particular case of the DDS one.

After that, we propose model parameter High-Resolution (HR) estimation algorithms, named PDDS-D[1] and PDDS-MC[2] and we explain why it is necessary to use HR methods in the audio transient modeling problem context. Finally, we show the efficiency of this approach on synthetic fast time-varying signals and on two typical audio transients.

II. Delayed sinusoidal models

A. PDDS model definition

In [18], we presented the M-order parametric DDS model. In this approach, every waveform 1-DDS possesses a delay parameter : $\{t_m\}_{1\leq m\leq M}$. Yet, in an audio modeling application, it is sufficient to consider a small number K of transient signals on a N-sample analysis such as $K \ll M$ (typically, $K < 3$ for $N = 512$ samples). We note k the index of the k-th transient signal and we fix

Rémy Boyer and Karim Abed-Meraim are with the Department of Signal and Image Processing, Ecole Nationale Suprieure des Tlcommunications (ENST), 46, rue Barrault, 75634 Paris Cedex 13, France. E-mail: [boyer, abed]@tsi.enst.fr.

[1]D stands for Deflation.

[2]MC stands for Multi-Channel.

$$M \triangleq \sum_{k=0}^{K} M_k \tag{1}$$

where M_k is the modeling partial order to represent the k-th transient signal with a support of $N_k = N - t_k$ samples. We denote $\{t_0, t_1, \ldots, t_{K+1}\}$ the delay parameter set with $t_0 = 0$, $t_{K+1} = N-1$, $t_k < t_{k+1}$, $0 \leq t_k \leq N-1$ and $B_k = t_{k+1} - t_k$. In relation with assumption **(A.1)**, we define the real M_k-PDDS model for $n = 0, \ldots, N-1$, by :

$$\hat{x}_k(n) \triangleq \sum_{m=1}^{M_k} a_{m,k} e^{d_{m,k}(n-t_k)} \,.\, \cos(\omega_{m,k}(n-t_k)+\phi_{m,k}) \,.\, \psi(n-t_k). \tag{2}$$

In the previous expression, $d_{m,k}$ is the (negative) damping factor, $\omega_{m,k}$ is the angular-frequency and $a_{m,k}$ and $\phi_{m,k}$ are respectively the m-th real amplitude and the m-th initial phase of the k-th M_k-PDDS model. The poles are defined by $z_{m,k} = e^{d_{m,k}+i\omega_{m,k}}$. Moreover, the Heaviside function $\psi(n)$ is defined by "1" for $0 \leq n \leq N-1$ and "0" otherwise. Note that there is a unique delay t_k for a set (sum) of M_k EDS waveforms (see figure 1).

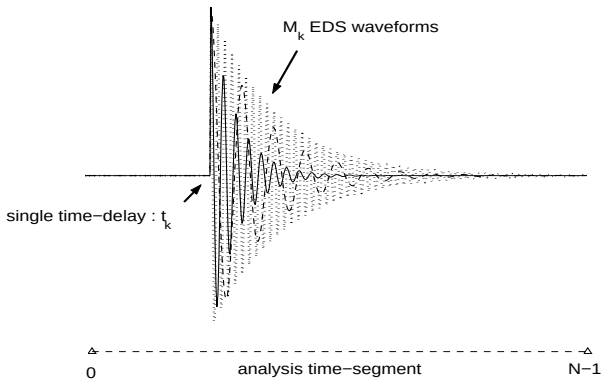

Fig. 1. M_k-PDDS model : one single time-delay for a set (sum) of M_k EDS waveforms.

Now, we can write the M-PDDS model expression as the sum of $(K+1)$ partial models, according to :

$$\hat{x}(n) \triangleq \sum_{k=0}^{K} \hat{x}_k(n). \tag{3}$$

B. Models Equivalence (ME)

If we assume that the signal $\hat{x}_k(n)$ is time shifted of the quantity "$+\,t_k$", we have, for $n = 0, \ldots, N_k - 1$:

$$\hat{x}_k(n+t_k) = \sum_{m=1}^{M_k} a_{m,k} e^{d_{m,k}n} \,.\, \cos(\omega_{m,k}n + \phi_{m,k}). \tag{4}$$

We recognize the expression of the real M_k-EDS model defined on a N_k-sample support. Moreover, we consider a second signal on the B_k-sample support $\{t_k, \ldots, t_{k+1}-1\}$, defined by $\hat{x}_k(n+t_k)$ where $0 \leq n \leq B_k - 1$. The latter can be seen as a truncated version of the signal of expression (4) by discarding the $N_k - B_k$ last samples. This operation is made in a view to eliminate the perturbation of the $(k+1)$-th transient signal. We conclude that if we have the knowledge of the delay t_k, then the time translation of the quantity "$+\,t_k$" of the M_k-PDDS model and the time support reduction (N_k to B_k) lead to consider an analysis by a M_k-EDS model on a B_k ($\leq N_k$) sample support. Once the model parameter estimation procedure is accomplished, we reconstruct the M_k-PDDS model by making the "inverse" operation, *i.e.*, a time support extension (B_k to N_k) and a translation of the quantity "$-\,t_k$". We define, in a similar way, the t_k-sample shifted audio signal by $x_k(n) = x(n+t_k)$ for $0 \leq n \leq B_k - 1$.

III. PDDS model parameters estimation

A. The need for High-Resolution (HR) method

Recalling that B_k is the effective analysis segment size of the k-th transient signal. This quantity can be quite small if the time-delay or the damping-factors are large. In these cases, it leads to a frequency resolution problem. Indeed, the Fourier resolution is of order $1/B_k$ for a B_k-sample segment. We can realize that the frequency resolution can be too coarse to make an efficient spectral analysis based on a Fourier-type method [11], [23]. Consequently, we use a HR method to jointly estimate the angular-frequencies and the damping-factors. These methods allow to overcome the Fourier resolution and perform well on very short time segments. More precisely, we will use the Kung's algorithm [25]. This method is based on the fundamental shift-invariance property of the signal basis.
Note that in the audio compression context, the total model order M is not estimated but fixed to reach a target bit-rate.

B. Delays estimation/detection

A transient signal can be seen as a very fast variation of the power of its envelop. So, in relation with assumption **(A.2)**, it seems natural to compute the envelop of the audio signal and to design a power transient detector based on the envelop variation. Consequently, we consider, here, a modified version of the detector, introduced in [26]. This modification consists of applying the detector on the smoothed signal envelop, rather than the audio signal. This improves slightly the detection/estimation performance.

B.1 Smoothed envelop

The smoothed envelop of the signal is computed by considering the median filtering of the modulus of the analytical signal $\nu_P(n)$. More precisely, we have :

$$\nu_P(n) \triangleq |\nu(n)| \,.\, f_P(n) \tag{5}$$

where the analytical signal is defined by $\nu(n) = x(n) + i\Psi_x(n)$, $\Psi_x(n)$ being the Hilbert transform of the audio signal and $f_P(n)$ is a median filter of length $2P$. Note that using a non-linear median filter allows to obtain a smoothed envelop of the signal, *i.e.*, without some awkward

oscillatory phenomena. On the other hand, this filter with short duration, typically $P = 5$ or less, keeps unchanged the global variation of the signal envelop.

B.2 The power transient detector

The second operation is to expose a transient detector which is based on the smoothed envelop power variations between two temporal hopping windows. The used formalism is the following :

$$\vartheta(n) \triangleq \frac{1}{J} \log \left(\frac{||\boldsymbol{\nu}_F(n)||_2^2}{||\boldsymbol{\nu}_B(n)||_2^2} \right) ||\boldsymbol{\nu}_F(n)||_2^2 \tag{6}$$

where $\boldsymbol{\nu}_B(n) = (\nu_P(n-J) \dots \nu_P(n-1))^T$ and $\boldsymbol{\nu}_F(n) = (\nu_P(n+1) \dots \nu_P(n+J))^T$. The vector $\boldsymbol{\nu}_B(n)$ (respectively $\boldsymbol{\nu}_F(n)$) represents the Backward (respectively Forward) time samples with respect to the analysis time n. J is the analysis depth. Note that the detector which was introduced in [26] works directly on the audio signal and not on the signal envelop. Consequently, our approach is an improve version of this detector.

B.3 Strategy of detection

In an audio transient detection application, two cases can occur.

B.3.a Single detection. The analysis segment is short enough to suppose that on its duration, there is in most one transient signal. This case is easily handled by maximizing the criterion $\vartheta(n)$, such as :

$$t_1 = \arg \max_{0 \leq n \leq N-1} \vartheta(n). \tag{7}$$

In our application and for an analysis duration between 128 and 512 samples, respectively 4 ms and 16 ms, we can reasonably suppose the presence of one transient signal at most during the analysis period.

B.3.b False and multiple detections. The first case is the false detection event, *i.e.*, the detector indicates the presence of a transient signal but visually, it is nothing. The second case is considered when the analysis segment is long enough to contain multiple transient signals. These two cases are handled with the introduction of a threshold s_t. Then, the estimated time-delays are the K local maximum values (larger than s_t) of the criterion $\vartheta(n)$. More precisely, the used strategy is the one introduced in [26].

C. PDDS-D algorithm : Deflation approach

This algorithm, summarized in Table I, is based on the following three procedures.

C.1 Partial orders allocation

In the introduction, we have mentioned that the most important feature for a percussive signal is its time waveform. Moreover, generally, the part before the onset has a weak power and is not very important for the naturalness of the transient. Inversely, the onset and the decreasing part have strong powers and are the most important parts of this kind of signal. Then, we choose to estimate the partial orders by the following empirical approach : small partial orders will be associated to low power signals since they do not need an accurate modeling. Inversely, higher power signals are associated to larger partial orders. Consequently, we introduce $\gamma \in \mathbb{R}$ according to :

$$M_k = \lceil \gamma \ . \ \varepsilon_k \rceil \tag{8}$$

where $\lceil . \rceil$ denotes the integer part and ε_k is the power of the B_k-sample audio signal $\boldsymbol{x}_k$, according to $\varepsilon_k = ||\boldsymbol{x}_k||_2^2 / B_k$. Afterwards, we fix :

$$\gamma = \frac{M+1}{\varepsilon_0 + \varepsilon_1 + \dots \varepsilon_K}. \tag{9}$$

C.2 Poles and complex amplitudes estimation

We begin by estimating the delays $\{t_k\}$ and the partial orders $\{M_k\}$ according to the previous methodologies. The PDDS-D algorithm principle is as follows : for each signal $\boldsymbol{x}_k$, we estimate the signal poles $\{z_{m,k}\}_{1 \leq m \leq M_k}$, according to the HR method and the complex amplitude parameters $\{\alpha_{m,k} = a_{m,k} e^{i\phi_{m,k}}\}_{1 \leq m \leq M_k}$ by resolution of the following linear least squares criterion :

$$\arg\min_{\boldsymbol{\alpha}_k} ||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k||_2 = \arg\min_{\boldsymbol{\alpha}_k} ||\boldsymbol{x}_k - \boldsymbol{Z}_k^{(B_k)} \boldsymbol{\alpha}_k||_2 \tag{10}$$

where :

$$\boldsymbol{Z}_k^{(B_k)} = \begin{bmatrix} \boldsymbol{\zeta}_{1,k} & \boldsymbol{\zeta}_{1,k}^* & \cdots & \boldsymbol{\zeta}_{M_k,k} & \boldsymbol{\zeta}_{M_k,k}^* \end{bmatrix} \tag{11}$$

is a $B_k \times (2M_k)$ Vandermonde matrix with $\boldsymbol{\zeta}_{m,k} = (1 \ z_{m,k} \dots z_{m,k}^{B_k-1})^T$. We, also, define $\boldsymbol{\alpha}_k = (\alpha_{1,k}, \alpha_{1,k}^*, \dots, \alpha_{M_k,k}, \alpha_{M_k,k}^*)^T$. The solution of criterion (10) is :

$$\boldsymbol{\alpha}_k = \boldsymbol{Z}_k^{(B_k)\dagger} \boldsymbol{x}_k \tag{12}$$

where $\dagger$ denotes the Moore-Penrose pseudo-inverse [27]. We, then, can synthesize the M_k-EDS model $\hat{x}_k(n + t_k)$. After that, we build the M_k-PDDS N-sample signal $\hat{x}_k(n)$ by a time support extension (B_k to N_k) and a "$-t_k$" shifting of the M_k-EDS model.

C.3 Deflation procedure

In a deflation procedure context, the algorithm begins by initializing the first residual signal $r_0(n) = x(n)$. At the k-th iteration and for the k-th residual signal $r_k(n)$, we estimate the M_k-EDS model : $\hat{r}_k(n + t_k)$ and we reconstruct the M_k-PDDS signal, $\hat{r}_k(n)$. Then, we add it to the synthesis signal $\tilde{x}_{k-1}(n)$. This operation is named the synthesis stage. And finally, we remove its contribution to the last residual signal $r_k(n)$ to compute the next residual signal $r_{k+1}(n)$. We summarize these two stages by :

$$\begin{array}{rcccc|l} \tilde{x}_k(n) & \triangleq & \tilde{x}_{k-1}(n) & + & \hat{r}_k(n) & \text{synthesis stage} \\ r_{k+1}(n) & \triangleq & r_k(n) & - & \hat{r}_k(n) & \text{analysis stage.} \end{array} \tag{13}$$

(1)	Delays estimation : $\{0, t_1, \ldots, t_K, N-1\}$
(2)	Partial orders allocation : $\{M_0, \ldots, M_K\}$
$k=0$	
(1)	**Initialization** : $r_0(n) = x(n)$
(2)	estimation of the M_0-EDS $\stackrel{ME}{\longrightarrow}$ M_0-PDDS : $\hat{r}_0(n)$
(3)	**synthesis** : $\tilde{x}_0(n) = \hat{r}_0(n)$
$k=1$	
(1)	**analysis** : $r_1(n) = r_0(n) - \hat{r}_0(n)$
(2)	estimation of the M_1-EDS $\stackrel{ME}{\longrightarrow}$ M_1-PDDS : $\hat{r}_1(n)$
(3)	**synthesis** : $\tilde{x}_1(n) = \tilde{x}_0(n) + \hat{r}_1(n)$
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$k=K$	
(1)	**analysis** : $r_K(n) = r_{K-1}(n) - \hat{r}_{K-1}(n)$
(2)	estimation of the M_K-EDS $\stackrel{ME}{\longrightarrow}$ M_K-PDDS : $\hat{r}_K(n)$
(3)	**synthesis** : $\tilde{x}_K(n) = \sum_{k=0}^{K} \hat{r}_k(n)$

TABLE I
PDDS-D ALGORITHM

D. PDDS-MC algorithm : "Multi-Channel" approach

We introduce, here, a second algorithm named PDDS-MC. All transients are treated jointly and thus only one single data matrix factorization is performed. Two versions of PDDS-MC algorithm are presented.

D.1 PDDS-MC1 algorithm

D.1.a First Hankel matrix factorization. It is possible to consider the analyzed segment as a set of "multi-channel" signals. In this approach, we estimate jointly the damping-factor and the angular-frequency parameters for the $(K+1)$ signals $\{\hat{x}_k(n+t_k), n=0,\ldots B_k-1\}_{0\leq k\leq K}$. We define the non-square $L_\nu \times L_k$ Hankel matrix $\mathcal{H}(\hat{\boldsymbol{x}}_k)$ such as $L_\nu + L_k = B_k$. We introduce the block-Hankel matrix according to :

$$\boldsymbol{H}(\hat{\boldsymbol{x}}) \stackrel{\triangle}{=} \left[\mathcal{H}(\hat{\boldsymbol{x}}_0) \quad \mathcal{H}(\hat{\boldsymbol{x}}_1) \quad \ldots \quad \mathcal{H}(\hat{\boldsymbol{x}}_K)\right]. \tag{14}$$

Its rank is $2M$ under condition that all the poles are different and without modeling noise. Every matrix $\mathcal{H}(\hat{\boldsymbol{x}}_k)$, represents the Hankel data matrix of the k-th channel of B_k samples size and verifies a factorization in a Vandermonde basis [25]. Consequently, $\boldsymbol{H}(\hat{\boldsymbol{x}})$ admits the following factorization :

$$\boldsymbol{H}(\hat{\boldsymbol{x}}) = \boldsymbol{\Theta} \;.\; \boldsymbol{\Lambda}_1 \tag{15}$$

where :

$$\boldsymbol{\Theta} = \left[\boldsymbol{Z}_0^{(L_\nu)} \quad \boldsymbol{Z}_1^{(L_\nu)} \quad \ldots \quad \boldsymbol{Z}_K^{(L_\nu)}\right] \tag{16}$$

and $\boldsymbol{\Lambda}_1$ is a non-singular matrix. We notice that factorization (15) highlights the row-shift invariance property of matrix $\boldsymbol{\Theta}$ which is a block-Vandermonde matrix. It is thus possible to use a HR method on $\boldsymbol{H}(\hat{\boldsymbol{x}})$ and to jointly determine the poles.

D.1.b Size of the block-Hankel data matrix. The choice of the parameters L_ν and $\{L_k\}$ is important since it influences the estimation performances of the PDDS-MC1 algorithm. In [6], it is shown that it is necessary to choose the row size L_ν of the data matrix such as $N/3 \leq L_\nu \leq 2N/3$. Moreover, we have $L_\nu + L_k = B_k$. Consequently, the L_k parameter has to satisfy $B_k - 2N/3 \leq L_k \leq B_k - N/3$. This condition implies a minimal bound of the channel size $B_k > N/3$. By considering the sum over k of the previous expression, we must have $K < 2$. In other words, to obtain maximum performance, the number of transient on the analysis segment must be 1. In the context of the transient audio modeling this is not a restrictive condition. In case of multiple transients, we fix $L_k = \min_k 2B_k/3$.

D.2 PDDS-MC2 algorithm : Second Hankel matrix factorization

Another approach is to consider the $(B_\nu - B_k)$-sample zero-padded signals $\hat{\boldsymbol{x}}_k^{(zp)}$ with $B_\nu = \max_k B_k$, according to $\hat{\boldsymbol{x}}_k^{(zp)} = [\hat{\boldsymbol{x}}_k^T \;\; \boldsymbol{0}_{B_\nu - B_k}^T]^T$. Based on the properties of the Hankel operator, we have :

$$\mathcal{H}\left(\sum_{k=0}^{K} \hat{\boldsymbol{x}}_k^{(zp)}\right) = \sum_{k=0}^{K} \mathcal{H}(\hat{\boldsymbol{x}}_k^{(zp)}) \approx \boldsymbol{\Theta} \;.\; \boldsymbol{\Lambda}_2 \tag{17}$$

where :

$$\boldsymbol{\Lambda}_2 = \left[\boldsymbol{Z}_0^{(L_\nu)}\boldsymbol{\Gamma}_0 \quad \boldsymbol{Z}_1^{(L_\nu)}\boldsymbol{\Gamma}_1 \quad \ldots \boldsymbol{Z}_K^{(L_\nu)}\boldsymbol{\Gamma}_K\right] \tag{18}$$

with $\boldsymbol{\Gamma}_k = \mathrm{diag}(\boldsymbol{\alpha}_k)$ and $B_\nu = 2L_\nu$ (square Hankel matrix). Due to the zero-padding, factorization (17) is only an approximation. However this approximation does not affect much the performance of the method. Note, we have to satisfy the constraint $4M \leq B_\nu$.

D.3 Poles processing

The $2M$ poles are estimated in the following manner :

$$\{z_{m,k}\} = \lambda_{2M}\left\{\boldsymbol{U}_{\downarrow}^{(2M)\dagger}\boldsymbol{U}_{\uparrow}^{(2M)}\right\}, \;\; \forall m, \; \forall k \tag{19}$$

where $\boldsymbol{U}^{(2M)}$ is the matrix containing the $2M$ left singular vectors of $\boldsymbol{H}(\hat{\boldsymbol{x}})$ or $\mathcal{H}(\hat{\boldsymbol{x}})$, $\lambda_{2M}\{.\}$ is the set of $2M$ eigenvalues and $\downarrow$ (respectively $\uparrow$) stands for deleting the bottom (respectively top) row. In presence of audio data (noisy data), we, simply, substitute $x_k(n)$ for $\hat{x}_k(n)$.

D.4 Filtering effect at the poles level

Let us notice that $\boldsymbol{\Theta}$, by definition, is (2M)-rank deficient matrix. If we assume that there exists $J(<M)$ identical poles which are simultaneously present in several channels then the rank of matrix $\boldsymbol{\Theta}$ decreases to $2(M-J)$. It follows when we decompose $\boldsymbol{H}(\hat{\boldsymbol{x}})$ or $\mathcal{H}(\hat{\boldsymbol{x}})$ through the rank-revealing factorization like the Singular Value Decomposition (SVD) [27], that we consider a $2(M-J)$ dimensional Signal basis. For real data matrix, $\boldsymbol{H}(\boldsymbol{x})$ or $\mathcal{H}(\boldsymbol{x})$, we have :

$$\text{rank}_\delta \left(\boldsymbol{U}^{(L_\nu)} \, . \, \text{diag}\{\sigma_1 \dots \sigma_{L_\nu}\} \right) = 2(M-J) \qquad (20)$$

where $\{\sigma_\ell\}_{1 \leq \ell \leq L_\nu}$ is the singular value set and $\text{rank}_\delta(.)$ stands for the numerical rank[3]. According to expression (20), we conclude that it is impossible to estimate several times the same pole, contrary to the PDDS-D algorithm. This property can be understood as a "filtering" property of the poles stemming from adjoining channels.

D.5 Pairing operation

For the two PDDS-MC methods, there is a pairing problem between the time-delays $\{t_k\}_{0 \leq k \leq K}$ and the couples $\{\omega_s, d_s\}_{1 \leq s \leq M}$. In other words, we have to associate the right time-delay to the right couple of angular-frequency and damping-factor. A simple way, to resolve this problem is, first, to compute a "collection" of waveforms $g_s(n) = e^{d_s n} \cos(\omega_s n + \phi_s)$ from the set of estimated couples $\{\omega_s, d_s\}$ and with $\phi_s = -\text{atan}(\alpha_2/\alpha_1)$ where $(\alpha_1 \ \alpha_2)^T = [\Re e(\boldsymbol{\zeta}_s) \ \Im m(\boldsymbol{\zeta}_s)]^\dagger \boldsymbol{x}_k$ and, second, to maximize over k the normalized correlation coefficient $\rho_{s,k}$ between each possible B_k-sample waveforms $\boldsymbol{g}_s$ and the audio signal $\boldsymbol{x}_k$. Then, for a given index s, we have :

$$\arg\max_k \rho_{s,k} \quad \text{where} \quad \rho_{s,k} \stackrel{\triangle}{=} \frac{|\langle \boldsymbol{g}_s \ , \ \boldsymbol{x}_k \rangle|}{||\boldsymbol{g}_s||_2 \ . \ ||\boldsymbol{x}_k||_2} \qquad (21)$$

and $\langle .,. \rangle$ denotes the scalar product. Note that from expression (21), we can, easily, deduce the modeling partial orders. Indeed, there exists a mapping between the set $1 \leq s \leq M$ and the set $0 \leq k \leq K$, then the modeling partial order $M_{k'}$ is the number of time that one component, among M, index by s is associated to the current index k', *i.e.*,

$$M_{k'} = \text{card} \left\{ 1 \leq s \leq M, \quad \arg \max_{0 \leq k \leq K} \rho_{s,k} = k' \right\} \qquad (22)$$

where card{.} denotes the cardinal.

D.6 Complex amplitudes estimation

The complex amplitudes are determined by solving the criterion :

$$\arg \min_{\{\boldsymbol{\alpha}_k\}} ||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2^2 \qquad (23)$$

where $\boldsymbol{x}$ is the N-sample audio signal and $\hat{\boldsymbol{x}}$ is the N-sample PDDS model of order M. By considering $\boldsymbol{J}_{t_k} = [\boldsymbol{0}_{N_k \times t_k} \ \boldsymbol{I}_{N_k}]^T$, a matrix which adds t_k rows of "0", we give the solution of the previous criterion :

$$\begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_K \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}_0^{(N)} & \boldsymbol{J}_{t_1} \boldsymbol{Z}_1^{(N_1)} & \dots & \boldsymbol{J}_{t_K} \boldsymbol{Z}_K^{(N_K)} \end{bmatrix}^\dagger \boldsymbol{x}. \qquad (24)$$

[3]defined in [27] by the number of $\sigma_\ell \geq \delta$ where δ is a fixed (positive) threshold.

IV. Fast-time varying signal modeling

A. Noisy synthetic signal

We consider a 100-sample noisy synthetic signal, according to :

$$\boldsymbol{x} = \hat{\boldsymbol{x}} + \sigma \boldsymbol{w} = \sum_{k=0}^{1} \boldsymbol{x}_k + \sigma \boldsymbol{w} \qquad (25)$$

where $\hat{\boldsymbol{x}}$ is a 100-sample PDDS signal. We add a white, Gaussian, unitary variance perturbation $\boldsymbol{w}$. Note that the variance of the random signal $\sigma \boldsymbol{w}$ is σ^2. The first part of the simulation deals with the sum of two 1-PDDS with separated time supports (see figure 2-a), *i.e.*, the first component has a sharp decreasing part (large damping-factor) in such a way that the second component is practically not disrupted. In this case, we will say that the two components are quasi-orthogonal, such as $\langle \hat{\boldsymbol{x}}_k, \hat{\boldsymbol{x}}_j \rangle \approx 0$ for $k \neq j$. The second part of the simulations is a study of the case where the components are non-orthogonal (see figure 2-b), *i.e.*, $\langle \hat{\boldsymbol{x}}_k, \hat{\boldsymbol{x}}_j \rangle \gg 0$.
The performance criterion is the Normalized Mean Square Error (MSE) evaluated for several Signal to Noise Ratios (SNR) using 100 "Monte-Carlo" trials. The MSE is defined by the ratio of the square difference between the true parameter value and its estimated value over the square value of the true parameter. Additionally, we define $\text{SNR}(\hat{\boldsymbol{x}}, \sigma \boldsymbol{w}) = 10 \log_{10}(||\hat{\boldsymbol{x}}||_2^2 / \sigma^2)$.

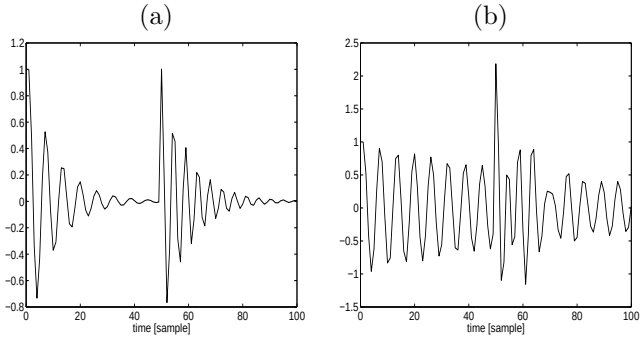

Fig. 2. Synthetic signal, (a) Quasi-orthogonal components, (b) Non-orthogonal components

A.1 Quasi-orthogonal case

We choose the following numerical values for the model parameters. $M_0 = M_1 = 1$, $\omega_{1,0} = 1$ rad, $\omega_{1,1} = 1.4$ rad, $d_{1,0} = d_{1,1} = -0.1$, $t_0 = 0$, $t_1 = 50$, $a_{1,0} = a_{1,1} = 1$ and $\phi_{1,0} = \phi_{1,1} = 0$. This signal is plotted on figure 2-a. On figures 3-a,b,c,d we expose the simulation results.
We can see that the PDDS-D and the PDDS-MC1 have very close performances. Only, negligible differences can be found between these two methods. For SNR higher than 5 dB, the PDDS-D algorithm is slightly more efficient. For SNR lower than 5 dB, this consideration is more mitigated. The PDDS-D presents better MSE values for the angular-frequency estimations but we observe a collapse of its performance for the damping-factor estimations at very

low SNR. The PDDS-MC2 is clearly less efficient than the two previous methods.

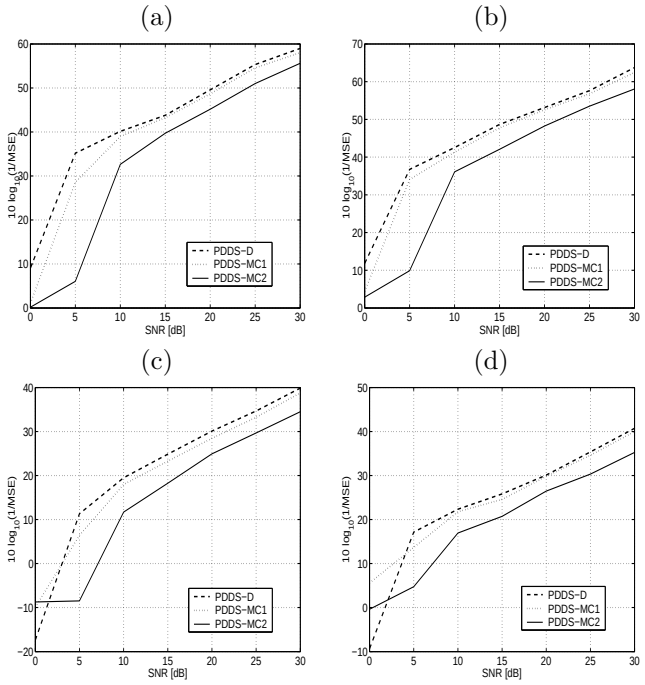

Fig. 3. Normalized Mean Square Error (MSE) Vs Signal to Noise Ratio (SNR) (a) $\omega_{1,0}$, (b) $\omega_{1,1}$, (c) $d_{1,0}$, (d) $d_{1,1}$.

A.2 Non-orthogonal case

We choose the following numerical values for the model parameters. $M_0 = M_1 = 1$, $\omega_{1,0} = 1$ rad, $\omega_{1,1} = 1.4$ rad, $d_{1,0} = -0.01$, $d_{1,1} = -0.1$, $t_0 = 0$, $t_1 = 50$, $a_{1,0} = 1$, $a_{1,1} = 3$ and $\phi_{1,0} = \phi_{1,1} = 0$. This signal is plotted on figure 2-b.

According to figures 4-a,b,c,d the PDDS-MC1 outperforms the two other methods, especially at low SNR (≤ 10 dB). This conclusion can be explained by considering the iterative scheme of the PDDS-D. Indeed, at low SNR, the estimation error at an early stage induces additional errors at the following ones. Inversely, for the PDDS-MC1, the "joint" character of the algorithm allows to keep high performance. In the non-orthogonal case, the PDDS-MC2 and the PDDS-D algorithms show similar performances.

B. Real audio signals

We test and compare the PDDS-D and PDDS-MC algorithms with the EDS approach on two 16 ms typical audio transient signals : triangle and castanet onsets. The sampling frequency is 32 kHz. Note that we have $4M$ parameters for the M-EDS model and $4M + K$ for the M-PDDS model. Moreover, according to our initial assumption, we have $K \ll M$. Consequently, in the context of parametric audio coding, the total number of model parameters is almost the same for the two models.

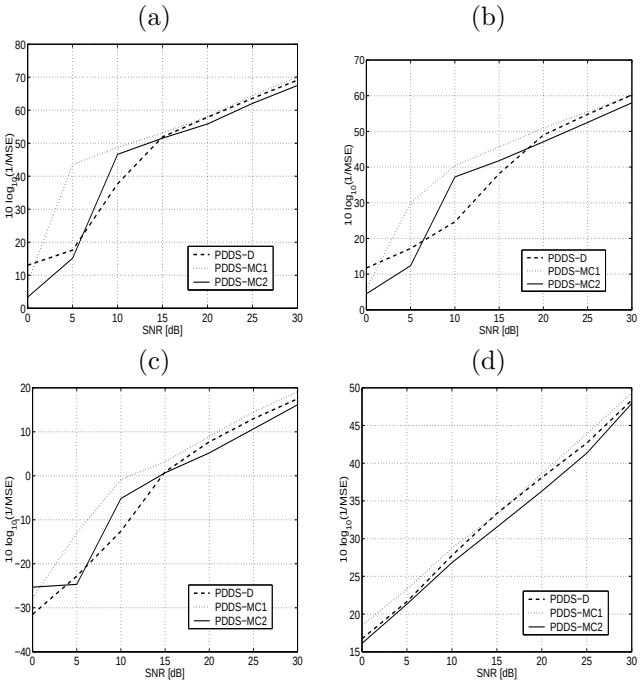

Fig. 4. Normalized Mean Square Error (MSE) Vs Signal to Noise Ratio (SNR) (a) $\omega_{1,0}$, (b) $\omega_{1,1}$, (c) $d_{1,0}$, (d) $d_{1,1}$.

B.1 First typical transient audio signal : triangle

B.1.a Time modeling. For this simulation, we choose a triangle onset since the attack has an extremely short duration (less than 50 samples). For this reason, this signal is extremely difficult to be efficiently modeled. We have represented the original waveform on figure 5-a. We can see on figure 5-b, the inefficiency of the EDS approach for a 28-order modeling. Note that the oscillating part of the signal is well represented but the dynamic onset is very low. This observation is confirmed by the SNR values in table II. Note that the SNR in the context of audio modeling is defined according to $\mathrm{SNR}(\boldsymbol{x}, \boldsymbol{r}) = 10\log_{10}(||\boldsymbol{x}||_2^2/||\boldsymbol{r}||_2^2)$ in dB where $\boldsymbol{r} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ is the residual audio signal. Inversely, the PDDS approach presents much better performances as we can see on figures 5-c,d,e and in table II.

	SNR [dB]	M_0 / M_1
PDDS-D	11.6	8 / 20
PDDS-MC1	9.8	9 / 19
PDDS-MC2	9.1	7 / 21
EDS	6.5	$M = 28$

TABLE II
SNR VALUES AND PARTIAL ORDERS ALLOCATION

We conclude that the PDDS approach outperforms the EDS approach and the PDDS-D algorithm shows the best SNR.

B.1.b "Time-frequency" analysis by filter-bank. Introducing a frequential aspect in the analysis, we use the polyphase 32-band pseudo-QMF filter-bank of MPEG1-

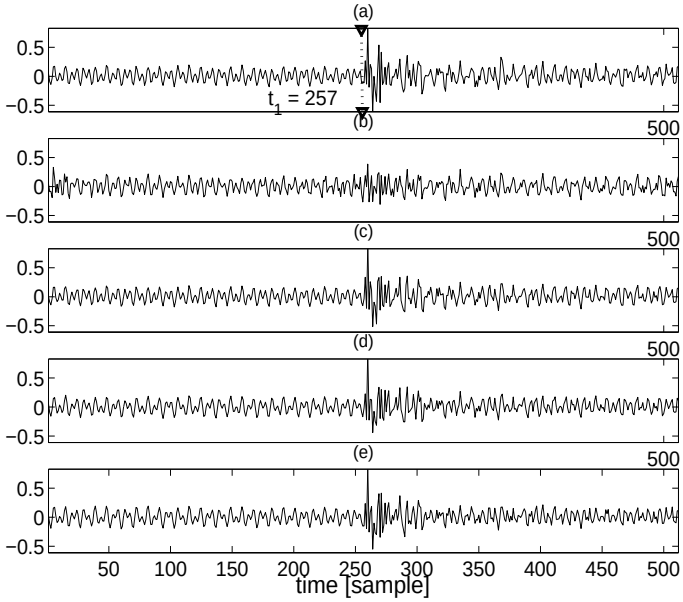

Fig. 5. (a) original triangle signal (normalized amplitude), (b) 28-EDS modeling, (c) 28-PDDS-D, (d) 28-PDDS-MC1, (e) 28-PDDS-MC2

audio [28] providing a uniform partition of the frequency axis. The bandwidth of each subband is 500 Hz with a 32 kHz sampling frequency. In each subband, we use the criterion SNR to characterize the "time-frequency" modeling performance of the considered model. This criterion is noted SNR_{TF}. According to figure 6 the PDDS approaches present, clearly, better SNR_{TF} than the EDS approach. We conclude that not only the onset is better represented (see figure 5) but also the whole audio signal is closer of the original signal, in the sense of the used criterion. This consideration is confirmed by the average SNR_{TF} over subbands, in table III.

PDDS-D	PDDS-MC1	PDDS-MC2	EDS
8.5	7.5	6.2	2.8

TABLE III
AVERAGE SNR_{TF} [DB] OVER SUBBANDS

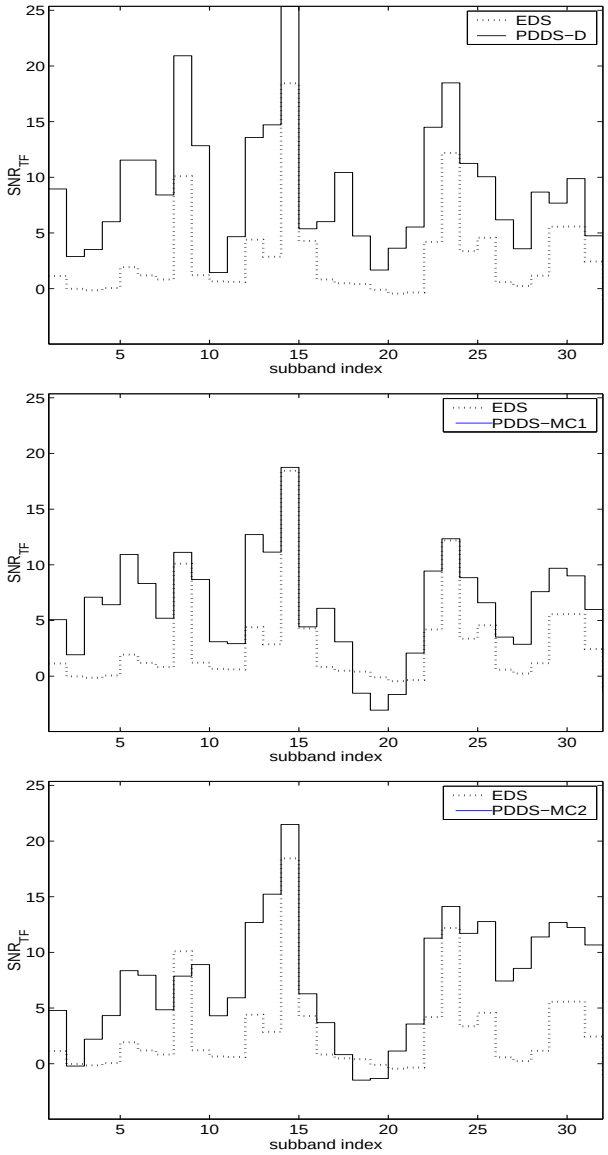

Fig. 6. PDDS Vs EDS in terms of SNR_{TF} criterion, (top) PDDS-D, (middle) PDDS-MC1, (bottom) PDDS-MC2

B.2 Second typical transient audio signal : castanets

In this part, we study the performances of the PDDS model and its robustness to a small error on the time-delay estimation.

B.2.a Time modeling. In this simulation, we fix the modeling orders to 20. On the top of figure 7, we have represented the original signal. On the middle of figure 7, we note the pre-echo[4] phenomenon and the weak dynamic onset for the EDS modeling. On the bottom of figure 7 and for the three methods, we can point out the total absence of pre-echo and the great reproduction of the onset dynamic. The PDDS model outperforms, clearly, the EDS approach.

[4]Additional energy before the onset.

B.2.b Perturbation of the estimated time-delay. Hereafter, we study the robustness of the PDDS-D and PDDS-MC algorithms to a perturbation Δ_t of the time-delay according to $t_1 + \Delta_t$ with $\Delta_t = \{-10, \ldots, 10\}$ on the castanet onset signal. The estimated time-delay t_1 for the castanet signal is 223 samples. Figure 8-b presents the partial order allocation for the three algorithms. On figure 8-a, we can see that the PDDS-D algorithm is the more robust algorithm, especially for time-delay under-estimation. The PDDS-MC2 is the less robust to the time-delay variation in the context of this simulation. The PDDS-MC1 shows intermediate robustness. Note that for the three methods, under-estimation is generally preferable to over-estimation.

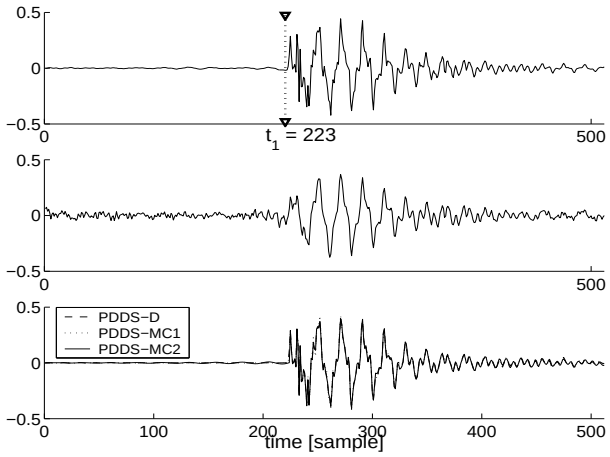

Fig. 7. (top) castanets onset (normalized amplitude), (middle) 20-EDS modeling, (bottom) 20-PDDS modeling for the three algorithms

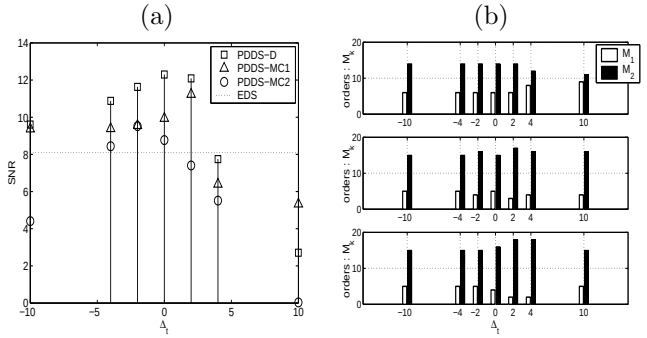

Fig. 8. (a) SNR with respect to the time-delay variation, (b) partial orders (top) PDDS-D, (middle) PDDS-MC1, (bottom) PDDS-MC2

V. Algorithmic complexity and choice of the algorithm

The complexity of the EDS algorithm can be evaluated to $O(NM^2)$ if we use an iterative processing of the SVD [29], [30]. The complexity of the PDDS-MC1 is similar to the EDS one. The computational cost of the PDDS-D algorithm can be evaluated to $O(\sum_k B_k M_k^2)$ and $O(B_\nu M^2)$ for the PDDS-MC2. Consequently, the PDDS-MC2 has the lowest computational complexity. Note that the cost of the time-delay and the partial order estimations are negligible.

From the simulation section, we conclude that the PDDS-D and PDDS-MC algorithms are well adapted to the transient audio modeling problem. Note that the allocation procedure for the PDDS-D algorithm is based on some empirical considerations on the "nature" of transient audio signal. Inversely, in the context of the PDDS-MC algorithms, the partial orders estimation is automatic since it is essentially a simple "re-allocation".

To conclude, we can say : for synthetic noisy signals, the PDDS-MC1 is the most efficient method since it presents similar performances than the PDDS-D algorithm in case of quasi-orthogonal components and superior performances (in particular for low SNRs) in case of non-orthogonal components.

For real audio signals and for the true time-delay estimation, the PDDS-D is the most attractive method since it has a moderate computational cost for slightly higher performance.

However, the PDDS-MC2 method can be chosen if the computational cost is the most important choice criterion, as often in the audio coding context.

In case of errors on the time delay estimation, we choose the PDDS-D method since this method presents the better trade-off between complexity, performances and robustness.

VI. Conclusion

In this paper, we have introduced an efficient non-stationary model for the transient compact representation problem. This model is an evolution of the DDS model introduced in the general context of signal modeling. This approach uses *a piori* information on percussive audio signal, *i.e.*, an audio transient signal can be seen as a sum of damped sinusoids with a single time-delay. This natural consideration leads to the proposed PDDS model and three high-resolution estimation methods. Finally, after having compiled the performance of the proposed methods on synthetic signals, we show that the PDDS approach outperforms the EDS approach on two typical transient audio signals. This conclusion is confirmed by intensive and informal listening tests.

References

[1] ISO-MPEG, *Call for proposals for new tools for audio coding*, ISO/IEC JTC1/SC29/WG11 MPEG2001/N3793, Jannuary 2001.

[2] B. Edler and H. Purnhagen, "Parametric audio coding", *Proc of the 5th International Conference on Signal Processing (ICSP 2000)*, Beijing, August 2000.

[3] A.C. Brinker, E.G.P. Schuijers, A.W.J. Oomens, "Parametric coding for high-quality audio", *Proc. of 112th AES convention*, Munich, Germany, May 2002.

[4] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio", *Proc of the IEEE*, Vol. 88, No 4, April 2000.

[5] R.J. McAulay and T.F. Quatieri, "Speech analysis & synthesis based on a sinusoidal representation", *IEEE Trans. on ASSP*, Vol. 34, No. 4, August 1986.

[6] A-J. Van Der Veen, ED. F. Deprettere and A. Lee Swindlehurst, "Subspace-Based Signal Analysis Using Singular Value Decomposition", *Proc. of the IEEE*, Vol. 81, No 9, September 1993.

[7] Y. Hua and T.K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise", *IEEE Trans. on Acoustic, Speech and Signal Pocessing*, Vol. 38 Issue: 5 , May 1990.

[8] H. Chen, *Subspace-Based Parameter Estimation of Exponentially Damped Sinusoids with Application to Nuclear Magnetic Resonance Spectroscopy Data*, PhD thesis, K.U.Leuven (Leuven, Belgium), May 1996.

[9] J. Nieuwenhuijse, R. Heusdens and E.F. Deprettere, "Robust Exponential Modeling of Audio Signal", *Proc. of Int. Conf. on Acoustic, Speech and Signal Processing*. Vol. 6 , 1998.

[10] J. Jensen, *Sinusoidal Models for Speech Signal Processing*, Ph.D. Thesis, Aalborg University, Denmark, August 2000.

[11] R. Boyer and J. Rosier, "Iterative method for Harmonic/Exponentially Damped Sinusoidal model", *Proc. of 5th Int. conf. on Digital Audio Effects (DAFx02)*, online article : `http://www.dafx.de/`, September 2002.

[12] K. Hermus, W. Verhelst, P. Wambacq and P. Lemmerling, "Total Least Squares based Subband Modelling for Scalable Speech Representations with Damped Sinusoids" , Proc. International

Conference on Spoken Language Processing, Vol. III, pages 1129-1132, Beijing, China, October 2000.

[13] R. Vafin, R. Heusdens, S. van de Par, and W. B. Kleijn, "Improved Modeling of Audio Signals by Modifying Transient Locations," *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01)*, New Paltz, NY, USA, 2001.

[14] P. Lemmerling, I. Dologlou and S. Van Huffel, "Speech Compression based on exact modeling and Structured Total Least Norm optimization", *Proc. of the IEEE Int. Conf. on Acoustic, Speech and Signal Processing*, May 1998.

[15] R. Boyer, S. Essid and N. Moreau, "Non-stationary signal parametric modeling techniques with an application to low bitrate audio coding" *Proc. of IEEE Int. Conf. Signal Processing*, August 2002.

[16] M. Goodwin and M. Vertterli, "Matching Pursuit and Atomic Signal Models Based on Recursive Filter Banks", *IEEE Trans. on Signal Pocessing*, Vol. 47, No. 7, July 1999.

[17] J. Jensen, R. Heusdens, "A Comparison of Sinusoidal Model Variants for Speech and Audio Representation", *Proc. of EUSIPCO*, 2002

[18] R. Boyer and K. Abed-Meraim, "Audio transients modeling by Damped & Delayed Sinusoids (DDS)", *Proc. of IEEE Int. Conf. on Acoustic, Speech and Signal Processing*, May 2002.

[19] R. Boyer, S. Essid and Nicolas Moreau, "Dynamic temporal segmentation in parametric non-stationary modeling for percussive musical signals", *IEEE Int. Conf. on Multimedia and Expo (ICME 02)*, August 2002.

[20] E. Zwicker and H. Fastl, *Psychoacoustics*, Springer-Verlag, Berlin, 1990.

[21] J. Laroche and J-L. Meillier, "Multichannel Excitation/Filter Modeling of Percussive Sounds with Application to the Piano", *IEEE Trans. on Speech and Audio Processing*, Vol. 2, No. 2, April 1994.

[22] P. Prandom, M. Goodwin and M. Vetterli, "Optimal Time Segmentation for Signal Modeling and Compression", *Proc. of the IEEE Int. Conf. on Acoustic, Speech and Signal Processing*, 1997.

[23] T. Verma and T. Meng, *A Perceptually Based Audio Signal Model with Application to Scalable Audio Compression*, PhD thesis, Stanford University, 1999.

[24] R. Boyer and S. Essid, "Transient modeling with a Frequency-Transform Subspace Algorithm and "Transient + Sinusoidal" scheme" *Proc. of IEEE Int. Conf. on Digital Signal Processing*, July 2002.

[25] S.Y. Kung, K.S. Arun and D.V. Baskar Rao, "State-space and singular-value decomposition-based approximation methods for harmonic retrieval problem", *J. Opt. Soc. Am.*, 73(12):1799-1811, December 1983.

[26] J. Kliewer and A. Mertins, "Audio Subband Coding With Improved Representation Of Transient Signal Segments", *Proc. of Europ. Signal Processing Conf.*, September 1998.

[27] G.H. Golub and C.F. Van Loan, *Matrix Computation*, North Oxford Academic, Oxford, second edition, 1983.

[28] K. Banderburg and G. Stoll, "ISO-MPEG-1 Audio : a generic standard for coding of high-quality digital audio", *JASA*, Vol. 42, October 1994.

[29] R. Boyer, "Fast algorithm and non-stationary model for high-resolution audio signal modeling", ENST internal report : `http:/www.tsi.enst.fr/~boyer/`, 2001.

[30] R. Badeau, R. Boyer and B. David, "EDS parametric modeling and tracking of audio signals", *Proc. of 5th Int. conf. on Digital Audio Effects (DAFx02)*, online article : `http://www.dafx.de/`, September 2002.