Modélisation de Signaux Transitoires Audio par Sinusoïdes Amorties et Retardées (SAR)

Rémy Boyer, Karim Abed-Meraim

ENST, Département Signal
46, rue Barrault, 75634, Paris Cedex 13
{boyer,abed}@tsi.enst.fr

Résumé

On propose, dans cet article, un nouveau modèle paramétrique basé sur une profonde évolution des modèles sinusoïdaux de McAulay et Quatieri [4]. Ce modèle prend en compte un paramètre d'atténuation et un paramètre de retard, ce qui le rend capable de modéliser efficacement des signaux fortement transitoires tels que les attaques de castagnettes et ceci sans créer de pré-echo[1] *et sans perte de dynamique au niveau de l'attaque. En parallèle, on développe un algorithme original de détermination des paramètres de modèle en utilisant une méthode haute-résolution suivie d'une analyse par sous-bande.*

Mots Clés

Modélisation audio, signaux transitoires, modèle sinusoïdaux, méthode haute-résolution, analyse en sous-bandes adaptées

1 Introduction

La représentation ou la modélisation d'un signal audio $s(n)$, $n = 0, \ldots N-1$ dans une optique de compression audio est un sujet actif depuis de nombreuses années. Un des modèles souvent utilisés pour atteindre ce but est le modèle Sinusoïdes Amorties Exponentiellement (SAE),

$$\hat{s}(n) = \sum_{m=1}^{M} a_m e^{d_m n} \cos\left(\omega_m n + \phi_m\right) \qquad (1)$$

où M est la dimension du sous-espace signal et $\{a_m, \phi_m, b_m, w_m\}$ sont les $4M$ paramètres d'amplitudes, de phases, d'amortissements, de pulsations. Ces modèles sont depuis longtemps étudiés dans la communauté des traiteurs de signaux. Cependant, leurs utilisations dans la compression du signal audio est assez récente [1], [2], [3]. Cette démarche s'inscrit dans une évolution logique des modèles sinusoïdaux introduits par McAulay et Quatieri [4] dans les années 1980. En effet, les modèles sinusoïdaux supposent que les paramètres de modélisation sont à variations lentes au regard de la durée N d'analyse. Or cette hypothèse est rarement vérifiée lorsque l'on traite des signaux audio très diverses comme la parole, la voix chantée ou la musique. Les modèles SAE en permettant aux termes d'amplitudes $\{a_m\}$ de varier exponentiellement avec le temps sont plus à même de représenter les signaux à fortes variations temporelles [1].
Cependant, il est bien connu que la modélisation de signaux transitoires par modèle SAE perd de son efficacité quand le signal cible est placé loin du début de la fenêtre d'analyse [1]. Plusieurs approches pour s'affranchir de ce problème ont vu le jour. Elles sont principalement de trois types, la première [5] se base sur l'utilisation d'une découpe irrégulière de l'axe des temps. La seconde [2], [6] estime directement le retard du transitoire en effectuant une modélisation sinusoïdale dans le domaine transformé de la DCT (Discret Cosine Transform), enfin la troisième [3] se base sur la construction et le stockage d'une famille de SAE indexée par un triplet : amortissements, pulsations et retards, sur laquelle est projeté le signal audio $s(n)$.

2 Modèle SAR

Une nouvelle approche possible est de modifier le modèle SAE afin de prendre en compte un paramètre de retard. On introduit la forme d'onde SAR (Sinusoïdes Amorties & Retardées) telle que

$$s_m(n) = a_m e^{i\phi_m} e^{(n-t_m)(d_m + i\omega_m)} u(n - t_m) \qquad (2)$$

où t_m est m-ème retard et $u(n)$ est l'échelon de Heaviside. Le modèle M-SAR réel d'ordre M est donné par

[1] Energie en "surplus" en amont de l'attaque.

$$\hat{s}(n) = \frac{1}{2}\left(\sum_{m=1}^{M} s_m(n) + s_m^*(n)\right) \quad (3)$$

Notons que l'ordre M du modèle sera fixé et non estimé. Dans cet article, on se propose, donc, d'explorer la représentation de signaux audio transitoires (attaques de castagnettes) par ce nouveau modèle paramétrique.

3 Algorithme proposé

On définit en préambule l'algorithme de détermination de M amplitudes et de M pulsations par la notation $\mathcal{A}_M(\{s(n)\}_{n=0}^{N-1}, \omega, d)$. Cet algorithme repose sur la propriété structurelle d'invariance par décalage de lignes (ou colonnes) de la matrice contenant une base du sous-espace signal associée aux $2M$ pôles $z_m = e^{d_m \pm i\omega_m}$ du signal $s(n)$. On pourra choisir par exemple l'algorithme 'Matrix Pencils' [9].

Notons que le critère non-linéaire global à résoudre est

$$\arg \min_{\omega,\phi,d,t,a} \sum_{n=0}^{N-1} |s(n) - \hat{s}(n)|^2 \quad (4)$$

Il serait irréaliste de vouloir résoudre ce problème d'optimisation conjointement. On développe alors la stratégie suivante

1. On cherche une première approximation des M pulsations sur le signal audio $s(n)$ telles que

$$\{\omega_m^{(1)}\} = \mathcal{A}_M\left(\{s(n)\}_{n=0}^{N-1}, \omega, d\right) \quad (5)$$

2. On construit le banc de filtres $\{h_m(n)\}$ où $h_m(n)$ est un filtre de réponse en fréquence passe-bande, de bande passante étroite et centré sur la pulsation $\omega_m^{(1)}$. On filtre le signal audio selon

$$s_m^h(n) = h_m(n) * s(n), \ \forall m \quad (6)$$

où $s_m^h(n)$ est la contribution du signal audio dans la direction $\omega_m^{(1)}$.

3. Dans chaque sous-bande indexée par m, on ré-estime la pulsation et on estime le paramètre d'atténuation, soit

$$\omega_m^{(2)}, d_m = \mathcal{A}_1\left(\{s_m^h(n)\}_{n=\rho_m}^{N-1}, \omega, d\right), \ \forall m \quad (7)$$

L'intérêt de ré-estimer la pulsation réside dans le fait qu'il est montré [8] que l'estimation de paramètres dans une sous-bande est plus performante que celle effectuée sur le signal en entier. Pour le terme d'atténuation, son estimation est un problème délicat et son calcul direct sur le signal est systématiquement biaisé quand celui-ci présente des retards[2]. Le terme $\rho_m = \arg\max_n |s_m^h(n)|$ est introduit afin de décaler l'instant initial de l'analyse pour faciliter l'estimation des paramètres d'atténuation. On le choisit, donc, logiquement dans le voisinage du maximum du signal filtré car on fait l'hypothèse que celui-ci est bien représentable par la forme d'onde SAR de l'équation (2).

4. A ce stade, on a la connaissance des M pulsations et atténuations, donc des pôles $\{z_m\}$. On résout le critère dans chaque sous-bande m

$$\arg\min_{\lambda,t} \sum_{n=0}^{N-1} \left|s_m^h(n) - \lambda z_m^{n-t} u(n-t)\right|^2 \quad (8)$$

Ce qui revient en ayant déterminé le λ optimal à résoudre

$$\arg\max_t \frac{\sum_{n=0}^{N-1} z_m^{*(n-t)} u(n-t) s_m^h(n)}{\sum_{n=0}^{N-1} \left|z_m^{n-t} u(n-t)\right|^2} \quad (9)$$

On résout ce critère par une simple énumération des valeurs possibles de t.

5. La dernière étape réside dans l'estimation des amplitudes complexe $\alpha_m = a_m e^{i\phi_m}$ selon le critère moindres carrés linéaire

$$\min_{\boldsymbol{\alpha}} \|\mathbf{s} - \mathbf{V}\boldsymbol{\alpha}\|_2^2 \quad (10)$$

où $\mathbf{s}$ est le signal audio sur N échantillons, $\boldsymbol{\alpha}$ est le vecteur composé des $2M$ termes $\{\frac{1}{2}a_m e^{\pm i\phi_m}\}$. La matrice $\mathbf{V}$ de dimension $N \times 2M$ possède alors la structure suivante

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_1^* & \dots & \mathbf{v}_M & \mathbf{v}_M^* \end{bmatrix} \quad (11)$$

avec $\mathbf{v}_m = [\mathbf{0}_{t_m} \ 1 \ z_m \dots z_m^{N-t_m-1}]^T$ où $\mathbf{0}_{t_m}$ est le vecteur ligne comportant t_m zéros. Enfin, on extrait de $\boldsymbol{\alpha} = \mathbf{V}^\dagger \mathbf{s} = (\mathbf{V}^H\mathbf{V})^{-1}\mathbf{V}^H\mathbf{s}$ les M amplitudes réelles $\{a_m\}$ et les M phases $\{\phi_m\}$.

4 Discussion

Nous présentons ici quelques remarques et commentaires sur la méthode proposée.

- L'estimation des paramètres d'atténuation est très sensible aux retards. La solution retenue dans ce travail a été de tronquer toute la partie du signal antérieure à l'instant où l'amplitude du signal est maximale (voir figure 1).

[2] Même si ces valeurs de retards sont faibles devant N.

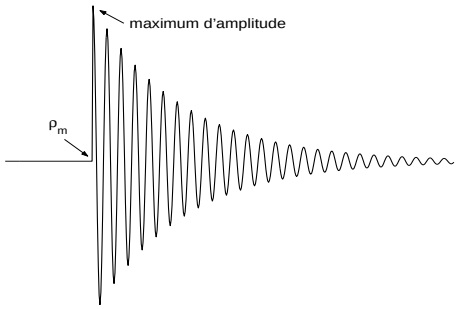

Figure 1: forme d'onde SAR et choix de ρ_m

Cette solution donne de bon résultats dans le cas où le modèle (2) est bien satisfait. Une solution probablement plus robuste au bruit de modélisation serait d'estimer conjointement le retard et l'atténuations dans chaque sous bande selon:

$$\arg\max_{t,d} \frac{\sum_{n=0}^{N-1} z_m^*(d)^{n-t} u(n-t) s_m^h(n)}{\sum_{n=0}^{N-1} |z_m(d)^{n-t} u(n-t)|^2}$$

où $z_m(d) = e^{i\omega_m + d}$.

- Notons qu'une erreur (même faible) sur l'estimation du retard induit une erreur forte sur l'estimation de la phase initiale du signal. En effet on a

$$ae^{i\phi} e^{(i\omega+d)(n-t)} = ae^{dk} e^{i(\phi+k\omega)} e^{(i\omega+d)(n-t-k)} \quad (12)$$

où k représente une erreur d'estimation sur le retard t. Cette erreur a un faible impact sur l'estimation de l'amplitude vu que $e^{dk} \approx 1$ pour k faible et $d \ll 1$ par contre l'erreur de phase (ici égale à $k\omega$) est importante. Toutefois, ces erreurs n'apparaissent pas (du fait qu'elles se compensent comme on le voit dans l'expression (12)) dans la modélisation finale du signal audio.

- La méthode 'Matrix Pencils' utilisée dans ce travail est adaptée au cas où le bruit est temporellement blanc. Le bruit de modélisation d'un signal audio est en général coloré. Il serait intéressant alors d'utiliser des méthodes d'estimation de pulsation et atténuation mieux adaptées à ce cas de figure, e.g. [7].

- Nous avons utilisé une méthode haute-résolution pour l'estimation de la pulsation et de l'atténuation dans chacune des sous-bandes afin de mieux combattre le résidue d'interférence des autres composantes. Cependant pour réduire la complexité on pourrait utiliser une méthode moins coûteuse telle que la FFT.

- Notons que l'utilisation du modèle SAR ne possédant que des valeurs de retards non nulles conduirait à ne pas modéliser la partie en amont de ces retards, il faut donc forcer à zéro un petit nombre de ces paramètres afin que l'ensemble du signal soit modélisé (dans notre simutation nous avons choisi d'en forcer 3). Une alternative serait de modéliser cette partie amont dans le signal résiduel.

5 Simulations sur un signal de castagnettes

Pour illustrer la capacité du modèle SAR à correctement modéliser des signaux très fortement transitoire, on choisit de comparer une modélisation par le modèle SAE à celle obtenue par le modèle SAR sur 512 échantillons de castagnettes, soit 16 ms. Pour le modèle SAE l'algorithme d'estimation des paramètres de pulsations et d'atténuations est un algorithme du type 'Matrix Pencils' [9]. Les paramètres d'amplitudes et de phases sont quant à eux calculés par résolution du critère moindres-carrés linéaire. On choisira le nombre de composantes pour le modèle SAE selon le rapport $M_{SAE} = \frac{5}{4} M_{SAR}$ afin de prendre en compte le paramètre de plus existant dans le modèle SAR. Sur les figures 2, 3 et 4 sont représentés les modélisations par les deux modèles selon différents instants initiaux d'analyse. On peut voir sur les figures 2-b, 3-b et 4-b que le modèle SAR ne génère pas de phénomène de pré-echo et modélise très correctement l'attaque de castagnettes. Les deux principaux défauts du modèle SAE sont donc corrigés par l'ajout de paramètres de retards. De plus, si on note D la distance entre l'instant initial d'analyse et le "début" du transitoire et que l'on fait croître ce paramètre, on observe, logiquement, sur les figures 2-c, 3-c et 4-c et sur le tableau 1 un effondrement des performances du modèle SAE alors que le modèle SAR garde un RSB proche de 12 dB.

	38-SAE	30-SAR
$D \sim 123$	12	12
$D \sim 204$	5	12
$D \sim 221$	2	11

Table 1: RSB en dB des signaux modélisés

6 Conclusion et perspectives

Dans ce travail, on propose dans un premier temps une évolution importante des modèles sinusoïdaux classiques [4]. Ce nouveau modèle est nommé SAR pour Sinusoïdes Amorties & Retardées. Dans un second temps, on expose une méthode d'estimation des paramètres SAR par une méthode originale, se basant

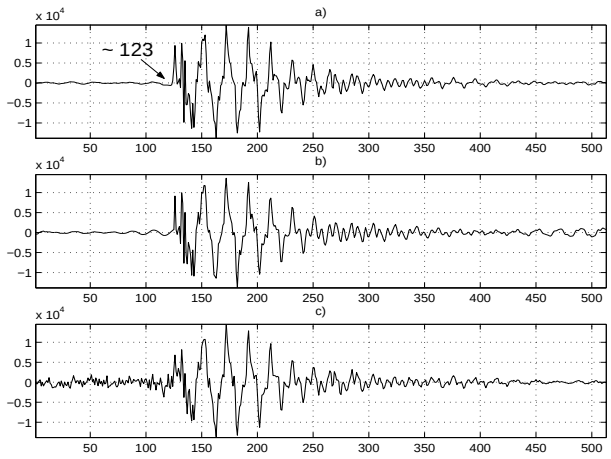

Figure 2: $D \sim 123$; a) signal original; b) 30-SAR; c) 38-SAE

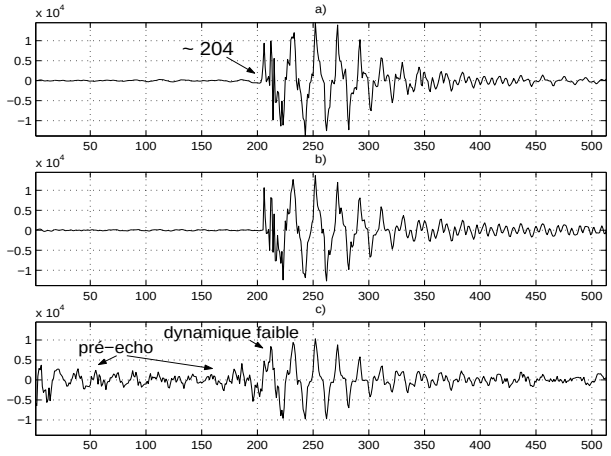

Figure 3: $D \sim 204$; a) signal original; b) 30-SAR; c) 38-SAE

sur un algorithme haute-résolution pour l'estimation des pulsations et une analyse par sous-bandes centrées autour des pulsations estimées du signal. On montre aussi sur un exemple de signal fortement transitoire (attaque de castagnettes) les limitations du modèle SAE et les qualités du modèle SAR. A savoir l'absence de pré-echo et une bonne reproduction de la dynamique de l'attaque du transitoire. Enfin, terminons en disant que les techniques développées, ici, ont une complexité algorithmique importante et que des méthodes de réduction de cette complexité sont à l'étude.

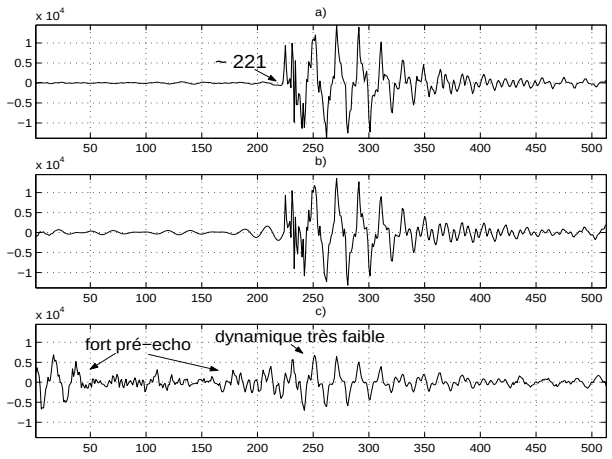

Figure 4: $D \sim 221$; a) signal original; b) 30-SAR; c) 38-SAE

References

[1] J. Nieuwenhuijse, R. Heusdens, E.F. Deprettere, *Robust Exponential Modeling of Audio Signal*, IEEE ICASSP. Volume: 6 , 1998

[2] R. Vafin, R. Heusdens, W. Bastiaan Kleijn, *Modifying Transients for Efficient Coding of Audio*, IEEE ICASSP, 2001

[3] M. Goodwin, *Matching Pursuit with Damped Sinusoids*, IEEE ICASSP, Volume: 3 , 1997

[4] R.J. McAulay, T.F. Quatieri, *Speech Analysis/Synthesis Based on a Sinusoidal Representation*, IEEE Trans. on ASSP, Vol 34, No 4, August 86

[5] P. Prandom, M. Goodwin, M. Vetterli, *Optimal Time Segmentation for Signal Modeling and Compression*, IEEE ICASSP, Volume: 3 , 1997

[6] T.S. Verma, T.H.Y. Meng, *An analysis/synthesis tool for transient signals that allows a flexible sines+transients+noise model for audio*, IEEE ICASSP, Volume: 6, 1998

[7] K. Abed-Meraim, A. Belouchrani, Y. Hua, A. Mansour, *Parameter Estimation of Exponentially Damped Sinusoids using Second Order Statistics*, EUSIPCO, 1998

[8] S. Rao, W.A. Pearlman, *On the superiority of coding and estimation from subbands*, Conf. on Information Sciences and Systems, March 1992

[9] Y. Hua, T.K. Sarkar, *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*, IEEE Trans. on ASSP, Volume: 38 Issue: 5 , May 1990