Exploration de techniques modernes de modélisation
adaptées à du codage audio bas-débit

Rémy BOYER Slim ESSID Nicolas MOREAU

Ecole Nationale Supérieure des Télécommunications

46, rue Barrault, 75634, Paris Cedex 13
boyer/essid/moreau@tsi.enst.fr

Résumé

On étudie des techniques modernes de représentation de signaux audio. Les limitations du modèle de Fourier nous poussent à étudier l'intérêt de l'utilisation de formes d'ondes mieux localisées à la fois en temps et en fréquence et les techniques s'y rattachant, en l'occurrence, les méthodes basées sur la décomposition en sous-espaces et par décompositions atomiques. On termine par la proposition d'un schéma de modélisation audio.

Mots Clef

Compression audio bas-débit, Modèle "Sinus + Bruit", Méthodes sous-espaces, Décompositions atomiques itératives, Modélisation des transitoires, CELP.

1 Introduction

Le domaine du codage audio a été très actif ces quinze dernières années. Récemment, un *call for proposals* dans le cadre de la normalisation MPEG4 a été annoncé dans l'optique de proposer un codeur fonctionnant à 24 kbits/s pour des signaux audio (parole et musique) de bande 20 Hz - 15 kHz délivrant une qualité acceptable. Notre contribution s'inscrit dans la philosophie des modèles sinusoïdaux [1]. On se propose, dans cet article, d'analyser les techniques de modélisation permettant de représenter aussi bien les signaux à caractère stationnaire que les signaux transitoires. Dans cette optique, deux approches ont été explorées. La première est basée sur la représentation paramétrique du signal dont les paramètres sont déterminés par une approche Haute-Résolution à l'aide de méthodes sous-espaces [2]. Ces techniques sont précédées par une procédure de séparation de sous-espaces [3]. La seconde approche exploite un algorithme itératif par projections successives du signal à modéliser sur les atomes d'un dictionnaire pré-défini [4] (Matching Pursuit MP et Orthogonal Matching Pursuit OMP).

2 Modélisation paramétrique et décomposition atomique

2.1 Expansion réelle sur une famille de formes d'ondes complexes

On part de l'hypothèse que le signal audio $s(n)$ peut être modélisé par la somme de deux composantes : une composante déterministe et une composante stochastique [1]. Soit, alors, $\mathcal{O}$ l'espace réel des observations tel que $dim\,(\mathcal{O}) = L$ et soit un signal audio de N échantillons, noté $\mathbf{s} = \{s(n)\}_{n=0}^{N-1}$ tel que $\mathbf{s} \in \mathcal{O}$. $\mathcal{O}$ peut être décomposé en deux sous-espaces : $\mathcal{S}$ le sous-espace signal de dimension $2M$ et $\mathcal{B}$ le sous-espace bruit tels que $\mathcal{O} = \mathcal{S} \oplus \mathcal{B}$, d'où l'écriture : $\mathbf{s} = \mathbf{s}_M + \mathbf{r}$ avec $\mathbf{s}_M \in \mathcal{S}$ et $\mathbf{r} \in \mathcal{B}$. Il vient, alors, $\mathbf{s}_M = \frac{1}{2}\sum_{m=1}^{M} \alpha_m \mathbf{g}_m + \alpha_m^* \mathbf{g}_m^* \in \mathbb{R}^{N\times 1}$ où $\mathbf{g}_m = \{g_m(n)e^{i\omega_m n}\}_{n=0}^{N-1}$, $\alpha_m = a_m e^{i\phi_m}$, w_m est la pulsation, a_m est l'amplitude réelle, ϕ_m est un terme de phase appartenant à l'ensemble $[0, 2\pi[$ et $(\mathbf{g}_1, \ldots, \mathbf{g}_M)$ est une base de $\mathcal{S}$ éventuellement orthonormale[1].

2.2 Choix du modèle de représentation

Limitations du modèle de Fourier. Les systèmes de modélisation de signaux audio par les méthodes "somme de sinus" se sont avérés efficaces pour modéliser des signaux harmoniques ou quasi-harmoniques, c'est à dire présentant des variations lentes au regard de la durée d'analyse [1]. Cependant, ces systèmes ne sont pas à même de représenter des signaux transitoires, typiquement les attaques ou évanouissements de sons, qui sont par nature localisés à la fois en temps et en fréquence. Il a paru donc intéressant de produire une représentation du signal à partir d'une famille de formes d'ondes plus "adaptées" permettant aussi l'étude de signaux à variation rapide.

Modèle SAE. On définit le modèle Sinusoïdal Amorti Exponentiellement d'ordre M (M-SAE) en posant $g_m(n) = e^{d_m n}$ où d_m est le m-ème coefficient

[1] $< \mathbf{g}_i, \mathbf{g}_j > = \delta_{i-j}$

d'amortissement réel. On notera que pour des $\{d_m\}$ nuls, le modèle est celui de Fourier.

Modèle de Gabor. Une famille d'atomes de Gabor peut être générée par dilatations, translations et modulations d'une fenêtre gaussienne $g(n) = e^{-\pi n^2}$. Pour tout paramètre d'échelle $s_m > 0$ et translation u_m, le modèle de Gabor est défini en posant $g_m(n) = g\left(\frac{n-u_m}{s_m}\right)$. En notant le triplet $\gamma_m = (s_m, u_m, \omega_m)$, on donne l'expression de l'atome normalisé de Gabor par $g_{\gamma_m}(n) = K_s g\left(\frac{n-u_m}{s_m}\right) e^{iw_m n}$ où K_s assure $\|\mathbf{g}_{\gamma_m}\|_2 = 1$.

2.3 Choix de la stratégie d'expansion

Deux modélisations ont été explorées : l'une est une modélisation paramétrique dont les paramètres sont déterminés à l'aide d'une approche Haute-Résolution au travers de méthodes sous-espaces [2] ; l'autre adopte une approche itérative par projections successives du signal à modéliser sur les atomes d'un dictionnaire déterminé [4].

Méthode Haute-Résolution (HR). Le modèle M-pôles admet la représentation matricielle $s_M(n) = \mathbf{1}_{2M}\mathbf{Z}^n\boldsymbol{\alpha}$ où on note $\mathbf{1}_{2M} = (1 \dots 1) \in \mathbb{R}^{1\times 2M}$, $\mathbf{Z} = diag\{e^{d_1+i\omega_1}, e^{d_1-i\omega_1}, \dots, e^{d_M+i\omega_M}, e^{d_M-i\omega_M}\}$, $\boldsymbol{\alpha} = \frac{1}{2}\left(a_1 e^{i\phi_1}\ a_1 e^{-i\phi_1} \dots a_M e^{i\phi_M}\ a_M e^{-i\phi_M}\right)^T$. On construit, alors, la matrice de Hankel $\mathbf{H} = \mathcal{H}_L(\mathbf{s}_M)$ à partir des données $s_M(n)$ et on exploite sa factorisation en deux matrices $\mathbf{O}_{(N-L)\times 2M}$ et $\mathbf{C}_{2M\times L}$ nommées respectivement matrice d'observabilité et matrice de commandabilité. De plus en définissant les deux opérateurs matriciels de décalage de ligne : $(.)_\uparrow$ la première ligne est supprimée et $(.)_\downarrow$ la dernière ligne est supprimée, on est en mesure de mettre en évidence la propriété d'invariance par décalage de ligne $\mathbf{O}_\downarrow\mathbf{Z} = \mathbf{O}_\uparrow$. L'expression précédente conduit alors à construire le produit[2] $\mathbf{O}_\downarrow^\dagger\mathbf{O}_\uparrow$ et d'en chercher une décomposition en valeurs propres. Soit une matrice $\mathbf{F}$ inversible et unitaire telle que $\mathbf{O}_\downarrow^\dagger\mathbf{O}_\uparrow = \mathbf{F}\mathbf{G}\mathbf{F}^H = \mathbf{Z}$. En se servant du fait que $\mathbf{F}$ est unitaire et de l'expression précédente, on peut dire que $\mathbf{G}$ et $\mathbf{Z}$ sont semblables et conservent donc le même spectre. Il est donc équivalent pour déterminer $\mathbf{Z}$ de diagonaliser $\mathbf{O}_\downarrow^\dagger\mathbf{O}_\uparrow$. Il importe maintenant de déterminer explicitement la matrice $\mathbf{O}$ puisque la factorisation de la matrice $\mathbf{H}$ n'est pas directement calculable en pratique. On choisira $\mathbf{O} = \mathbf{U}\mathbf{T}_{2M} = [\mathbf{u}_1 \dots \mathbf{u}_{2M}]$ avec $\mathbf{T}_{2M}$ une matrice de sélection des $2M$ premières colonnes de $\mathbf{U}$, la matrice de vecteurs singuliers à gauche. On en conclut que $Im(\mathbf{O}\mathbf{T}_{2M}) = Im([\mathbf{u}_1 \dots \mathbf{u}_{2M}]) = \mathcal{S}$ et donc que $(\mathbf{u}_1, \dots \mathbf{u}_{2M})$ est une base orthonormale. Pour des signaux réels, la troncature de la SVD n'est pas toujours évidente. On exploite donc l'algorithme Composite Property Mapping (CPM) introduit dans la référence [3] ici utilisé dans le but de "séparer" $\mathcal{S}$ de $\mathcal{B}$. Pour ce faire, on définit les opérateurs de troncature selon $\mathcal{T}_{2M}(\mathbf{H}) = \sum_{m=1}^{2M}\lambda_l\mathbf{u}_l\mathbf{v}_l^H$ et de moyennage sur les anti-diagonales $\mathcal{M}(\mathbf{H}) = \mathbf{h}$ où h_l est le résultat du moyennage sur la l-ème anti-diagonale de $\mathbf{H}$, soit $\mathbf{H}_k = [\mathcal{H}_L \circ \mathcal{M} \circ \mathcal{T}_{2M}]^k(\mathbf{H})$. L'étape suivante réside dans la détermination des amplitudes réelles et des phases en résolvant le problème classique moindres carrés $\min_{\boldsymbol{\alpha}} \|\mathbf{s} - \mathbf{V}\boldsymbol{\alpha}\|_2^2$ avec $\mathbf{V}$ la matrice de Vandermonde de dimension $N \times 2M$ de terme général $\{e^{(d_m \pm i\omega_m)n}\}$. Ce problème admet pour solution $\mathbf{a} = \mathbf{V}^\dagger\mathbf{s}$.

Discussion sur la méthode HR. Les méthodes Haute-Résolution ou basées sur la décomposition en sous-espaces ont cet avantage de se libérer de la limitation de la résolution de Fourier et sont capables de déterminer les valeurs numériques des paramètres de modèle avec une grande précision. Cette caractéristique peut être utile lorsque pour certains sons complexes, deux raies spectrales sont séparées par une distance inférieure à la résolution spectrale qui est de l'ordre de f_e/NHz. De même pour des situations où il existe des "glissements" de fréquences, cette méthode peut être efficace. Le deuxième avantage de l'utilisation de cette méthode est sa capacité à estimer correctement les paramètres de modèle sur un faible nombre d'échantillons. L'utilisation de fenêtres très courtes est alors envisageable sans dégradation majeure des performances d'estimation. Il est alors possible d'avoir une bonne résolution temporelle. Notons qu'avec une analyse de Fourier classique, la taille de la fenêtre conditionne la résolution fréquentielle et une analyse sur fenêtre courte est alors difficilement réalisable.

Poursuite adaptative orthogonale (MP et OMP). Le Matching Pursuit est un algorithme itératif introduit par Mallat et Zhang [4] qui réalise une décomposition du signal sur un ensemble d'atomes choisis dans un dictionnaire de taille $N \times D$ surcomplet ou redondant ($D \gg N$) tel que $\mathbf{G} = \{\mathbf{g}_\gamma\}_{\gamma\in\Gamma}$ où Γ est l'ensemble indexant les atomes du dictionnaire. Le principe a auparavant été utilisé en codage de la parole, dans les codeurs CELP, pour la sélection des vecteurs d'excitation [6]. On procède par approximations successives de $\mathbf{s}$, par projections orthogonales sur des "molécules diatomiques" obtenues à partir de $\mathbf{G}$. On effectue, en fait, des décompositions de signaux réels à l'aide de dictionnaires formés d'atomes complexes afin d'éviter de discrétiser un paramètre de phase supplémentaire et d'augmenter ainsi les dimensions du dictionnaire [5]. Cela est rendu possible grâce à une poursuite en sous-espaces conjugués (MPsec). Il s'agit de trouver, à chaque itération du MP, un sous-espace de dimension 2 engendré par un atome et son conjugué. A la m-ème itération, on cherche donc à trouver "une

[2] $()^\dagger$ note la pseudo-inverse

molécule" matérialisée par une matrice $(\mathbf{G}_{\gamma_m})_{(N\times 2)}$ dont les 2 colonnes sont formées d'un vecteur du dictionnaire et de son conjugué, qui minimise la norme du résiduel $\mathbf{r}_{m+1} = \mathbf{r}_m - \mathbf{G}_{\gamma_m}\boldsymbol{\alpha}_m$, où $\boldsymbol{\alpha}_m$ est un vecteur de 2 coefficients de corrélation et $\mathbf{G}_{\gamma_m}$ est choisi tel que $\mathbf{G}_{\gamma_m} = \arg\max_{\gamma\in\Gamma} |< \mathbf{G}_\gamma, \mathbf{r}_m >|$. La contrainte d'orthogonalité $< \mathbf{r}_m - \mathbf{G}_{\gamma_m}\boldsymbol{\alpha}_m, \mathbf{G}_{\gamma_m} >= 0$ implique une solution pour les poids $\boldsymbol{\alpha}_m = \mathbf{G}^{\dagger}_{\gamma_m}\mathbf{r}_m$. On a alors[3] $\mathbf{r}_{m+1} = \mathbf{r}_m - 2\Re(\alpha_m(1)\mathbf{g}_{\gamma_m})$. Avec cette démarche, on obtient des décompositions du signal de la forme $\mathbf{s}_M = 2\sum_{m=1}^{M}\Re(\alpha_m(1)\mathbf{g}_m)$. Une variante peut être considérée : le Matching Pursuit Orthogonal [4] [4] qui permet d'assurer que le résiduel $\mathbf{r}_m$ est orthogonal à tous les $m-1$ vecteurs du dictionnaire déjà sélectionnés dans le modèle et doit ainsi assurer une convergence rapide de l'algorithme. En fait, le MP nécessite une infinité d'itérations pour reconstruire $\mathbf{s}$ parfaitement, alors que l'OMP permet de s'assurer que la poursuite cesse après un nombre fini d'étapes [4]. L'algorithme initial est modifié comme suit : on orthonormalise à l'itération m le vecteur $\mathbf{g}_{\gamma_m}$ par rapport aux $m-1$ vecteurs $\mathbf{g}_{\gamma_1}, \ldots, \mathbf{g}_{\gamma_{m-1}}$ déjà sélectionnés, par le procédé de Gram-Schmidt, soit $\mathbf{g}^{\perp}_{\gamma_m} = \frac{\mathbf{g}_{\gamma_m} - \mathbf{P}_{\mathcal{V}_{m-1}}\mathbf{g}_{\gamma_m}}{\|\mathbf{g}_{\gamma_m} - \mathbf{P}_{\mathcal{V}_{m-1}}\mathbf{g}_{\gamma_m}\|_2}$ où $\mathbf{P}_{\mathcal{V}_{m-1}}$ est le projecteur orthogonal sur le sous-espace $\mathcal{V}_{m-1} = vect\{\mathbf{g}_{\gamma_1}, \ldots, \mathbf{g}_{\gamma_{m-1}}\}$ et l'on reprojette $\mathbf{r}_m$ sur $\mathbf{g}^{\perp}_{\gamma_m}$.

Discussion sur les méthodes MP et OMP. L'algorithme MP est un algorithme général de décomposition d'un problème d'optimisation de dimension M en M problèmes d'optimisation de dimension 1. Cette approche est donc sous-optimale par essence mais présente l'intérêt de ne faire aucune hypothèse, *a priori*, sur le signal étudié et permet d'utiliser une grande variété de formes d'onde pour composer le dictionnaire. On pourra citer en plus des formes d'ondes utilisées dans cet article, les Ondelettes ou les Chirps. Cependant, le pas de discrétisation des paramètres dans le dictionnaire conditionne ses performances. On peut souligner la capacité de cet algorithme à corriger aux itérations suivantes les artéfacts de modélisation créés à une itération courante. Cette propriété procure au MP une bonne robustesse lors de la modélisation de signaux de natures très diverses. Notons enfin que l'utilisation de l'OMP permet d'accélérer la vitesse de convergence de la norme du résiduel [4] mais ne se justifie que pour des dictionnaires sur-complets [5] puisque le fait d'avoir des atomes s'exprimant comme combinaison linéaire d'autres atomes du même dictionnaire implique une perte d'orthogonalité au sein de cette famille et nécessite donc lors de l'extraction de la base du sous-espace signal une ré-orthogonalisation de celle-ci.

[3] Le détail des calculs sera trouvé dans la réference [5]

[4] Là encore le principe a été utilisé en codage de parole [6]

[5] On définit un tel dictionnaire par la construction d'une famille de vecteurs génératrice mais non-libre.

3 Simulations sur signaux réels

Dans cette partie, on s'attache à comparer les performances des méthodes HR, MP et OMP sur des formes d'onde de Fourier, SAE et de Gabor dans le contexte de modélisation de signal. Le critère de comparaison utilisé est alors simplement le Rapport Signal à Bruit (RSB). On présente les mesures de RSB de modélisation sur trois types de signal : signal de parole (harmonique + transitoires doux) sur figure.2-a, signal de castagnettes (purement transitoire) sur figure.2-b et signal de clochettes (harmoniques + transitoires) sur figure.2-c. De plus, il nous parait important de comparer ces méthodes selon un critère de débit. Pour chacune des simulations, on fixe les paramètres de chaque modèle tels que $\rho^{(HR)} \simeq \rho^{(MP)} \simeq \rho^{(OMP)}$ où ρ est le débit estimé pour chacune des méthodes.
Sur les simulations, on observe que le MP et l'OMP présentent une bonne robustesse vis à vis du caractère non-stationnaire des signaux du type castagnettes (figure.2-e) et clochettes (figure.2-f). La méthode HR-Fourier est à peu près au niveau de MP et OMP sur les parties stationnaires (figure.2-d et 2-f) mais marque le pas nettement sur les parties transitoires (figure.2-e et 2-f). Les figures.2-g, 2-h et 2-i mettent en évidence les limitations de HR-Fourier et montrent que HR-SAE lui est supérieur en terme de RSB. On observe sur les figures.2-k et 2-l que les méthodes HR-SAE et OMP-Fourier sont faiblement performantes sur les signaux à caractère fortement transitoire comme les castagnettes ou l'attaque de clochette. Par contre, OMP-Gabor fournit de bons résultats sur ces signaux. Enfin, sur les figures.2-j, 2-k et 2-l, on constate que la méthode HR-SAE se comporte mieux que OMP-Fourier. Ce qui confirme que HR-SAE est bien adaptée aux signaux à caractère oscillant ou faiblement transitoire.

4 Conclusion et perspectives

A la lumière des simulations effectuées, on propose un schéma de modélisation sur la figure.1. Le signal audio est décomposé en trois composantes traitées successivement : transitoire, harmonique et stochastique. On ajoute pour le traitement des transitoires un pré et post-traitement par transformation DCT afin que OMP-Gabor ne modélise essentiellement que les parties transitoires du signal. Le dernier résiduel est modélisé par une technique CELP. L'analyse-synthèse est effectuée à l'aide de la méthode OLA à faible recouvrement. Les premiers résultats d'écoutes sont encourageants puisque nous avons pu constater, grâce à des tests informels, que la majorité des signaux modélisés

demeurent de qualité acceptable, c'est à dire que l'on ne relève aucune dégradation significative.

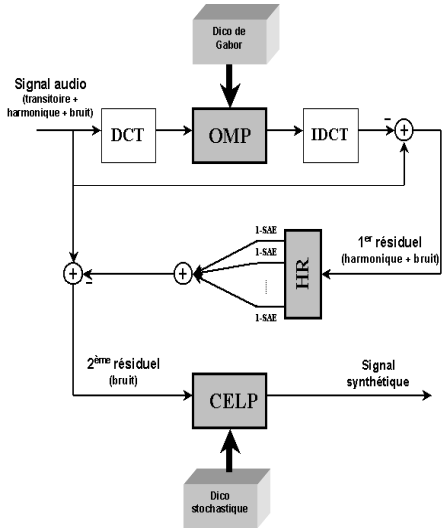

Figure 1: Schéma de modélisation

References

[1] X. Serra, J. Smith III, *Spectral Modeling Synthesis : A Sound System Based on a Deterministic plus Stochastic Decomposition*, Computer Music Journal, Vol 14, No 4, Winter 1990.

[2] S.Y. Kung, K.S. Arun, D.V. Bhaskar Rao, *State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem*, J. Opt. Soc. Am., Vol 73, No 12, December 1983.

[3] J.A. Cadzow, Signal Enhancement - *A Composite Property Mapping Algorithm*, IEEE Trans. on ASSP, Vol 36, No 1, January 1988 .

[4] S. Mallat, *Une exploration des signaux en ondelettes*, Editions Polytechniques, Novembre 2000.

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

[6] N. Moreau, P. Dymarski, *Selection of Excitation Vectores for the CELP Coders*, IEEE Trans on SP, Vol 2, No 1, January 1994

Annexe

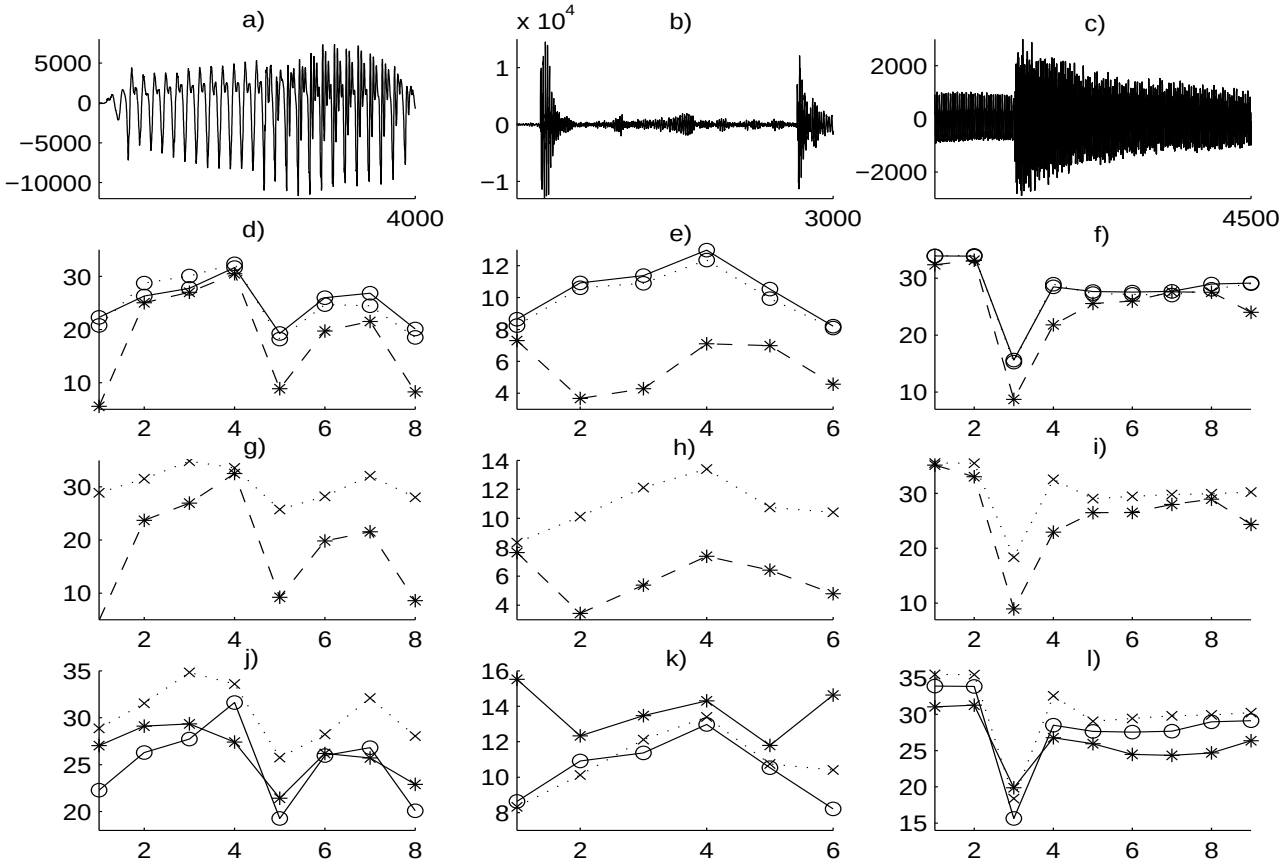

Figure 2: RSB par trame sur signaux tests
HR-Fourier *(*- -)*; HR-SAE *(×..)*
MP-Fourier *(o..)* ; OMP-Fourier *(o–)* ; OMP-Gabor *(*–)*