Estimation robuste de l'enveloppe spectrale d'un signal harmonique bruité

M. Oudot, O. Cappé, E. Moulines

ENST dpt. Signal / CNRS-URA 820
46 rue Barrault, 75634 Paris Cedex 13
oudot,cappe,moulines@sig.enst.fr

RÉSUMÉ

Les modèles sinusoïdaux de parole nécessitent l'estimation d'une enveloppe spectrale reliant les différents harmoniques. Une nouvelle méthode est développée ici ; elle repose sur un critère de vraisemblance pénalisée obtenu à partir du comportement statistique des carrés des amplitudes des sinusoïdes, estimés en présence de bruit de mesure. Un critère composite est développé afin de traiter les trames mixtes, dans lesquelles le signal utile de parole est composé de sinusoïdes et de bruit.

ABSTRACT

When sinusoidal speech models are used for coding purpose, it is necessary to design a smooth spectral envelope that connects the estimated harmonics. In this paper, a new method is introduced for parametric modeling of the spectral envelope in the presence of measurement noise. The proposed method is based on a penalized likelihood criterion which can be extended to the case of "mixed" (sinusoidal and noise) models.

1 Introduction

On considère que le signal analysé s'écrit sous la forme

$$s(t) = \sum_{k=1}^{K} \left[a_k \cos \omega_k t + b_k \sin \omega_k t \right] + n(t) \qquad (1)$$

où $n(t)$ est un bruit stationnaire de densité spectrale de puissance $\Pi_n(\omega)$, et les pulsations ω_k sont reliées harmoniquement. Dans le cadre du traitement de la parole, ce modèle est réaliste dès lors que l'on se restreint à des durées relativement brèves de l'ordre de quelques dizaines de millisecondes. Cette modélisation de type "sinusoïdes + bruit" est à la base de nombreuses applications tant en synthèse [7] qu'en codage [5].

La physiologie de la production de la parole permet de postuler l'existence d'une enveloppe spectrale qui relie les amplitudes des différents harmoniques du signal (dans les sons voisés au moins). L'estimation de cette enveloppe constitue une tâche particulière pour laquelle seule l'information présente aux fréquences des harmoniques peut être considérée comme significative. Il s'agit en fait d'un problème d'estimation spectrale "discret" qui a été pour la première fois exposé sous cette forme par El Jaroudi et Makhoul [2].

La méthode présentée dans [1] constitue une avancée dans la mesure où elle formalise l'attente intuitive d'une enveloppe spectrale "régulière" en introduisant une fonctionnelle de régularisation. Toutefois, cette méthode ignore totalement la présence du bruit $n(t)$: elle conduit à donner la même importance à tous les harmoniques quel que soit leur niveau. Ce comportement contredit l'intuition statistique qui voudrait que les harmoniques fortement bruités, dont l'amplitude ne peut être estimée que de façon peu précise, ne soient pas mis sur le même plan que les harmoniques de très fort niveau. Nous présentons dans une première partie les fondements théoriques de la méthode d'estimation proposée (dite Penalized Likelihood Estimation ou PLE). La seconde partie est consacrée aux résultats obtenus, à ordre fixe, pour une paramétrisation cepstrale ou auto-régressive. Enfin la dernière partie permet de mieux comprendre l'effet de la régularisation en étudiant le comportement du PLE en fonction des valeurs du facteur de régularisation.

Ce travail a été financé par le Centre National d'Étude des Télécommunications (CNET) sous le contrat 95-PE-7697

2 Vraisemblance asymptotique

2.1 Cas d'une trame voisée

Supposons que nous disposons d'un signal correspondant au modèle (1). Dans le cas d'une trame voisée, le bruit $n(t)$ représente le bruit de mesure ainsi que l'inaptitude du modèle à représenter le signal original en n'utilisant que des sinusoïdes. Soient a_k et b_k les amplitudes des composantes en phase et en quadrature de la k-ième sinusoïde, estimées à partir de la Transformée de Fourier Discrète (DFT).

Lorsque la taille T de la fenêtre d'analyse est suffisamment grande, les valeurs estimées $\hat{a}_k$ et $\hat{b}_k$ sont asymptotiquement normales de moyennes respectives a_k et b_k, de variance n_k et de plus, statistiquement indépendantes.

$$\begin{bmatrix} \hat{a}_k \\ \hat{b}_k \end{bmatrix} \sim AN \left(\begin{bmatrix} a_k \\ b_k \end{bmatrix}, \begin{bmatrix} n_k/2 & 0 \\ 0 & n_k/2 \end{bmatrix} \right) \qquad (2)$$

avec [1]

$$n_k = \frac{4G_{\bar{w}}}{T}\Pi_n(\omega)$$

L'amplitude carrée estimée de la $k$ième sinusoïde $x_k = (\hat{a}_k)^2 + (\hat{b}_k)^2$ suit alors une loi du χ^2 non centrée à deux degrés de liberté [4] dont la densité de probabilité est donnée par :

$$p(x_k) = \frac{1}{n_k}\ \exp[-\frac{s_k + x_k}{n_k}]\ I_0\left(2\sqrt{\frac{s_k x_k}{n_k^2}}\right)$$

où $s_k = a_k^2 + b_k^2$ est la vraie valeur de l'amplitude carrée de la $k$ième sinusoïde et $I_0()$ la fonction de Bessel modifiée de la première espèce, d'ordre $\nu = 0$. L'expression de ces fonctions pour un ordre ν est rappelée ci-dessous :

$$I_\nu(y) = \frac{1}{\pi}\int_0^\pi e^{-ycos\theta}cos(\nu\theta)d\theta$$

En pratique les fonctions de Bessel peuvent être calculées à partir de leur développement en série.

Grâce à l'hypothèse d'indépendance des observations, l'opposé de la log-vraisemblance associée aux K amplitudes estimées $\mathcal{L}(x_{1,\dots,K}|S)$ s'écrit :

$$\mathcal{L}(x_{1,\dots,K}|S) = \sum_{k=1}^{K}\left[\log n_k + \frac{s_k + x_k}{n_k} - \log I_0\left(2\sqrt{\frac{s_k x_k}{n_k^2}}\right)\right] \quad (3)$$

S désigne l'enveloppe spectrale au carré et $s_k = S(\omega_k)$. Puisqu'aucune hypothèse n'a encore été avancée concernant la paramétrisation, n'importe laquelle peut être utilisée (AR, cepstre, etc.).

2.2 Cas d'une trame mixte

Une trame mixte contient à la fois des composantes voisées et non-voisées. Cette trame est modélisable par un modèle à deux bandes, réduction pratique du modèle multi-bandes [3]. Nous supposons donc que le signal est voisé jusqu'à la fréquence dite de coupure F_c et qu'il est bruité au-delà. Le critère PLE développé ci-dessus ne s'appliquant qu'à la région voisée du spectre, il nous faut développer un critère mixte.

Soit y_k la valeur du périodogramme correspondant à la fréquence $\lambda_k = 2\pi k/T$. Si T est suffisamment grand alors y_k suit une loi de χ^2 centrée à deux degrés de liberté donnée par :

$$p(y_k) = \frac{1}{\Pi_y(\lambda_k)}\exp\left[-\frac{y_k}{\Pi_y(\lambda_k)}\right]$$

Ainsi, si S est la représentation paramétrique de la densité spectrale originale Π_y et $s_k = S(\lambda_k)$ alors la log-vraisemblance associée à tout le périodogramme est donnée par [6] :

$$W(y_0, \dots, y_{T/2}|S) = \sum_{k=0}^{T/2}\left[\log s_k + y_k/s_k\right]$$

[1] G_h est une constante de normalisation qui ne dépend que des caractéristiques de la fenêtre de pondération $\bar{w}$ utilisée.

ce qui constitue l'approximation de Whittle, plus connue dans le domaine de la parole sous le nom de critère d'Itakura Saito.

En pratique, nous utilisons la vraisemblance associée aux points du périodogramme situés au-delà de F_c, c'est-à-dire commençant à l'indice K'.

Il est important de noter que si le signal de parole est contaminé par un bruit ambiant $\Pi_n(\lambda)$, le critère peut être modifié de façon à ne modéliser que le bruit utile $\Pi_s(\lambda)$.

$$\Pi_y(\lambda) = \Pi_s(\lambda) + \Pi_n(\lambda)$$

il vient alors, avec $n_k = N(\lambda_k)$ où N est la représentation de Π_n :

$$W(y_{K'}, \dots, y_{T/2}|S) = \sum_{k=K'}^{T/2}\left[\log(s_k + n_k) + \frac{y_k}{s_k + n_k}\right]$$

2.3 Régularisation

Bien souvent, la minimisation directe de l'opposé de la log-vraisemblance (3) conduit à des solutions physiquement inacceptables en tant qu'enveloppes spectrales. Il est donc nécessaire de contraindre le comportement de l'enveloppe à estimer, grâce à une fonction de régularisation. Notre critère dit CPLE (Composite Penalized Likelihood Estimation) devient :

$$L(x_1, \dots, x_K|S) + W(y_{K'}, \dots, y_{T/2}|S) + \lambda R(S) \quad (4)$$

λ est appelé facteur de régularisation. Typiquement, une fonction de la forme suivante convient.

$$\mathcal{R}(S) = \int_{-\pi}^{\pi}\left[\frac{d^r \log S(\omega)}{d\omega}\right]^2 d\omega^r$$

Une telle régularisation est particulièrement adaptée à une paramétrisation cepstrale, pour des raisons théorique (choix de $\log S(\omega)$ donc d'une distance dans le domaine log-spectral, perceptivement plus significative) et pratique (elle s'exprime directement en fonction des coefficients cepstraux) [1] [6].

3 Résultats

3.1 Paramétrisation cepstrale

Nous nous intéressons ici au cas d'une paramétrisation cepstrale d'ordre élevé ($p = 44$). La procédure de minimisation du critère est initialisée en utilisant la méthode du cepstre discret. Celle-ci estime les coefficients cepstraux par une méthode de moindres carrés sur les amplitudes exprimées en échelle log et utilise, elle aussi, la fonctionnelle de régularisation présentée ci-dessus [1]. La figure 1 illustre le comportement du critère de vraisemblance appliqué à une trame test composée de sinusoïdes et de bruit blanc de puissance connue. On remarque que lorsque les sinusoïdes sont peu entachées de bruit, elles sont parfaitement reliées par l'enveloppe, alors qu'un effet de sous-estimation affecte celles qui sont noyées dans le bruit. les observations correspondant à des sinusoïdes émergeant clairement du bruit sont donc très bien estimées alors que les autres, ne pouvant être précisément déterminées, ne sont pas surestimées, comme dans les procédures classiques d'interpolation utilisant une méthode d'extraction de pics. Il

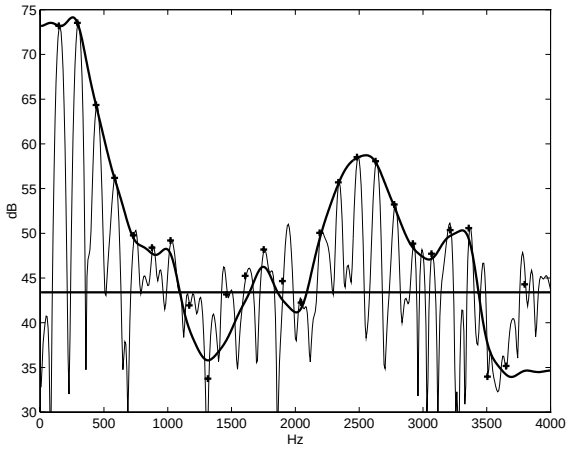

FIG. 1 — Enveloppe cepstrale optimisée d'un signal test "harmoniques + bruit de mesure". $\lambda = 5e^{-4}$. Les croix correspondent aux sinusoïdes. Le niveau de bruit est indiqué par un trait horizontal.

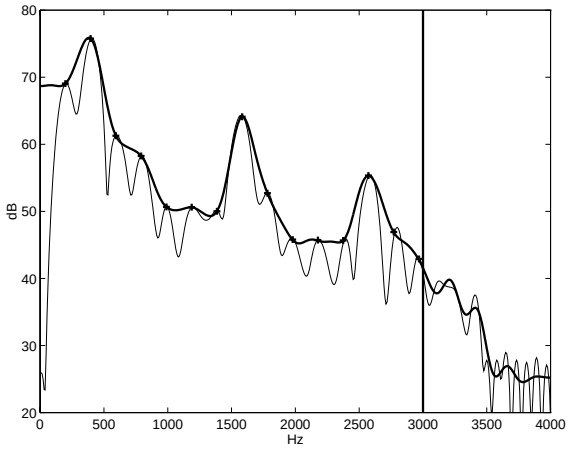

FIG. 2 — Enveloppe cepstrale composite d'un signal réel. $\lambda = 5e^{-4}$. $F_c = 3000Hz$ représentée par un trait vertical. Les croix correspondent aux sinusoïdes.

est important de noter que c'est le critère de régularisation qui impose la forme de l'enveloppe dans les régions bruitées. La figure 2 illustre la modélisation d'un signal réel obtenue par CPLE. Le comportement basses fréquences est régi par le critère PLE commenté précédemment et l'on observe que dans les hautes fréquences le critère composite estime la densité spectrale du bruit.

3.2 Paramétrisation Auto-Régressive

Lorsqu'une paramétrisation d'ordre élevé n'est pas possible pratiquement (cas du codage à bas débit), le choix d'une paramétrisation auto-régressive est tentant. Nous nous plaçons ici dans le contexte d'un codeur à débit faible utilisant une enveloppe AR d'ordre 16. la procédure d'optimisation est initialisée avec des paramètres auto-régressifs issus d'une analyse fréquentielle. Le choix des paramètres initiaux est dans ce cas important, car ce type de paramétrisation conduit à des problèmes de minima locaux du critère. Il est en outre à noter que les calculs sont beaucoup plus lourds que dans le cas d'une paramétrisation cepstrale. La figure 3 met en évidence

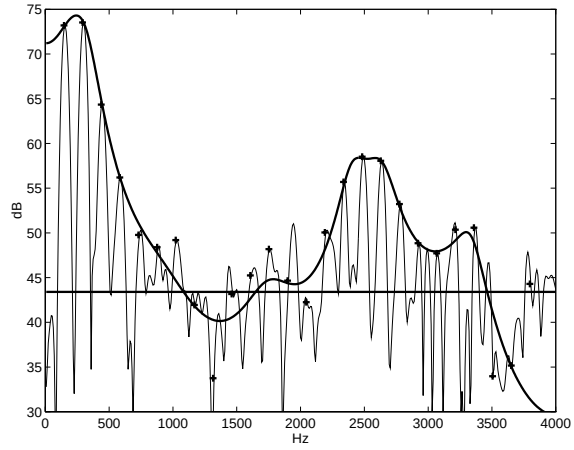

FIG. 3 — Enveloppe auto-régressive d'un signal test "harmoniques + bruit de mesure". $\lambda = 5e^{-4}$. Les croix correspondent aux sinusoïdes. Le niveau de bruit est indiqué par un trait horizontal.

le comportement du PLE ainsi que l'effet de l'ordre de la paramétrisation, pour les mêmes données que la figure 1. La

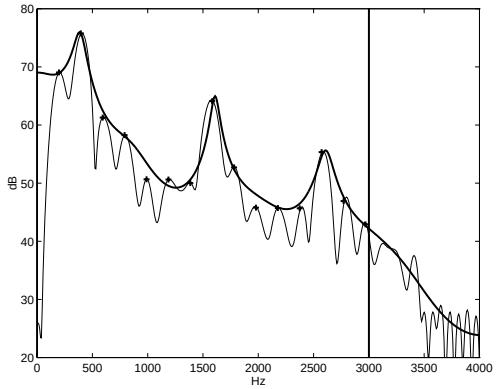

FIG. 4 — Enveloppe auto-régressive composite d'un signal réel. $\lambda = 1e^{-3}$. Les croix correspondent aux sinusoïdes. $F_c = 3000Hz$ représentée par un trait vertical.

figure 4 montre le résultat du critère mixte sur la même trame de signal réel que la figure 2.

3.3 Discussion

La valeur du facteur de régularisation choisie dans les expériences précédentes ainsi que l'effet de la fonction de régularisation nécessitent quelques explications. La figure 5 met en évidence la nécessité d'une fonction de régularisation : la courbe obtenue pour $\lambda = 1e^{-6}$ a une allure inacceptable. Le facteur de régularisation doit être bien choisi, sinon l'on risque aussi de perdre la précision recherchée (enveloppe obtenue pour $\lambda = 1$), précision qui dépend déjà de l'ordre de la paramétrisation. Comment choisir ce facteur ? La figure 6 a été obtenue, sur des signaux tests, en calculant l'erreur quadratique moyenne entre les enveloppes originale et estimée

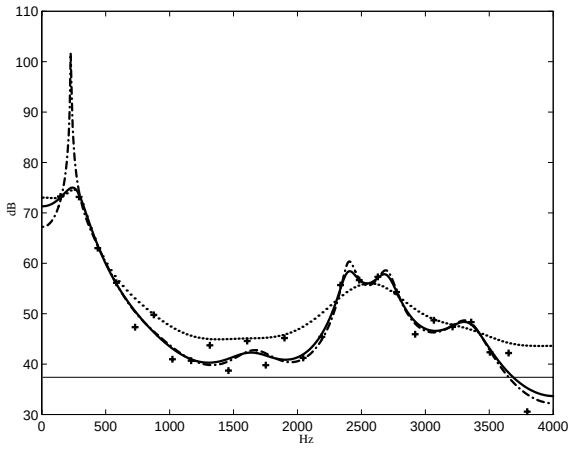

FIG. 5 — Enveloppes auto-régressives obtenues pour $\lambda = 1e^{-6}$ (traits mixtes) $\lambda = 1e^{-3}$ (trait continu) et $\lambda = 1$ (pointillés). Les croix correspondent aux sinusoïdes. Le niveau de bruit est indiqué par un trait horizontal.

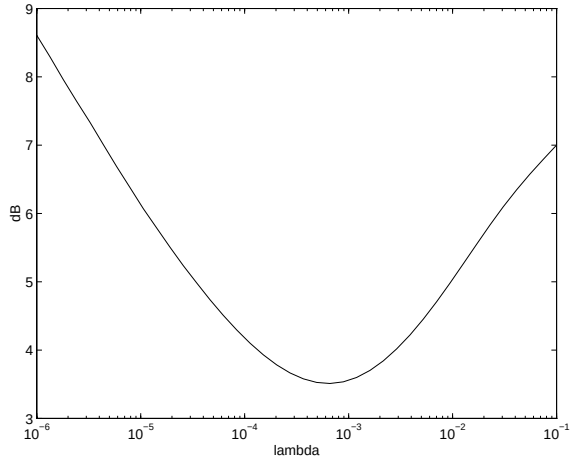

FIG. 6 — Erreur en dB en fonction de la valeur de λ entre l'enveloppe original et l'enveloppe estimée par cepstre discret (traits pointillés) et par PLE (trait continu).

(cepstrale), exprimées sur une échelle log, pour différentes valeurs du facteur de régularisation λ. On constate qu'il existe une valeur minimale de l'erreur ce qui signifie, qu'en présence de bruit de mesure, une valeur bien choisie du facteur de régularisation permet à la procédure d'estimation de mieux approcher l'enveloppe originale. l'existence de cette valeur minimale n'est pas étonnante : on constate en effet que lorsque le facteur λ est trop petit, l'enveloppe prend, en moyenne, des valeurs très faibles dans les zones bruitées et peut présente de larges oscillations dans les autres zones ; lorsqu'au contraire les valeurs du facteur de régularisation sont trop grandes, l'enveloppe est plus régulière mais l'on perd en précision d'interpolation des sinusoïdes. Différentes expériences ont montré que l'on obtenait ce type de courbes, pour diverses combinaisons d'enveloppes, de fréquences fondamentales et de Rapports Signal sur Bruit (RSB). Pour une enveloppe et un RSB donnés, on constate que la valeur du minimum de l'erreur est la même quelle que soit la fondamentale ; si l'on diminue le RSB, la position du minimum se décale vers la gauche mais la valeur de l'erreur est moins sensible au réglage de λ. Ces expériences montrent que la valeur $\lambda = 5e^{-4}$ convient à tout type d'enveloppe, même si elle ne correspond pas forcément à la valeur optimale.

4 Conclusion

Une nouvelle méthode d'estimation de l'enveloppe spectrale, reposant sur une critère de maximum de vraisemblance, a été présentée. Elle peut être étendue à un critère composite permettant de traiter les trames mixtes. Aucun choix de paramétrisation a priori n'ayant été fait, ce critère s'adapte à tout type de paramétrisation, mais le choix cepstral reste le plus adapté du point de vue calculatoire. Une grande robustesse au bruit a été démontrée et l'effet de la régularisation a été mis en évidence.

Références

[1] O. Cappé and E. Moulines. Regularization techniques for discrete cepstrum estimation. *IEEE Signal Processing Letters*, 3(4) :100–102, apr 1996.

[2] A. El-Jaroudi and J. Makhoul. Discrete all pole modeling. *IEEE Trans. Signal Processing*, 39(2) :411–423, February 1991.

[3] D. W. Griffin and J. S. lim. Multiband excitation vocoder. *IEEE Trans. Acoust., Speech, Signal Processing*, 36(8) :1223–1235, august 1988.

[4] N. L. Johnson and S. Kotz. *Continuous Univariate Distributions*, volume 2. Wiley-Interscience, 1970.

[5] R. J. Mc Aulay and T. F. Quatieri. Sinusoidal coding. In W.B. Kleijn and K.K. Paliwal, editors, *Speech Coding and Synthesis*, pages 123–176. elsevier, 1995.

[6] Y. Pawitan and F. O'Sullivan. Non parametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association*, 89(426) :600–610, june 1994.

[7] Y. Stylianou, J. Laroche, and E. Moulines. High-quality speech modification based on a harmonic + noise model. In *Proc. EUROSPEECH*, Madrid, sep 1995.