

Convergence locale de filtres en cascade. Application à l'annulation d'écho non linéaire

Alexandre GUÉRIN^{1,2}, Régine LE BOUQUIN-JEANNÈS¹, Gérard FAUCON¹

¹Laboratoire de Traitement du Signal et de l'Image, Université de Rennes 1, Bâtiment 22, Campus de Beaulieu, 35042 Rennes cedex, France

²Alcatel Business Systems, Micro-Systems and Signal Processing Department (ABS-MSD), 32 Avenue Kléber, 92707 Colombes cedex, France

alexandre.guerin@univ-rennes1.fr, regine.lebouquin@univ-rennes1.fr, gerard.faucon@univ-rennes1.fr

Résumé –

Cet article décrit un algorithme d'annulation d'écho acoustique non linéaire traitant plus particulièrement des distorsions dues au haut-parleur. La structure proposée est composée de deux modules adaptatifs organisés en cascade. L'étude théorique de la convergence de systèmes en cascade est réalisée à l'aide de filtres de Wiener et met en évidence le phénomène de **convergence locale**. Une stratégie d'adaptation est proposée permettant de stabiliser le système : des résultats sur signaux réels sont donnés qui montrent la stabilité de l'algorithme et sa supériorité comparé à un filtrage adaptatif linéaire.

Abstract –

This paper describes a nonlinear acoustic echo cancellation algorithm, dealing especially with the loudspeaker distortions. The proposed system is decomposed into two distinct modules organized in a cascaded structure. The theoretical study of the convergence of such systems by the means of Wiener filters is derived leading to the **local convergence** phenomenon. An adaptation strategy is proposed to stabilize the adaptive system : results on real recorded data are given showing the stability of the proposed algorithm and its superiority over a classical linear filter.

1 Introduction

Les terminaux GSM sont, pour la plupart, dotés de la fonctionnalité mains-libres « intégré ». Il en résulte non seulement une augmentation notable de l'écho acoustique, mais également des distorsions de cet écho dues au comportement non linéaire du haut-parleur aux tensions limites d'utilisation : l'emploi d'algorithmes linéaires ne suffit plus pour assurer une annulation d'écho convenable. La figure 1 illustre ce constat en comparant les performances en termes d'ERLE (Echo Return Loss Enhancement) de l'algorithme linéaire NLMS (Normalized-Least Means Squares), pour le même signal envoyé dans un cas à la tension maximale préconisée par le constructeur, dans l'autre à la tension maximale moins 15dB : la perte de performances est très importante, même pour des distorsions non audibles sur le signal de parole.

Afin de prendre en considération cet aspect non linéaire, Fermo propose un annuleur d'écho à base de filtres de Volterra d'ordre 2 [1]. Dans [2], les auteurs modélisent la saturation de l'amplificateur de puissance du haut-parleur par un polynôme. On se propose ici d'étudier un algorithme d'annulation d'écho non linéaire basé sur une modélisation par deux filtres adaptatifs en cascade du couple haut-parleur/canal acoustique : le haut-parleur est modélisé par un filtre de Volterra d'ordre 3, tandis que le canal acoustique est représenté par un filtre linéaire à réponse impulsionnelle (R.I.) finie. L'étude portera particulièrement sur l'adaptation conjointe des filtres en cascade, pour lesquels on mettra en évidence le problème de convergence locale.

2 Modèle du système

Du fait de l'élasticité de la membrane, les haut-parleurs présentent des propriétés de mémoire, aussi bien dans leur partie non linéaire (lorsque la tension d'entrée est proche de la tension maximale) que dans leur partie linéaire : de par leur comportement harmonique, la base des filtres de Volterra est particulièrement adaptée à leur modélisation. La figure 2 décrit le système d'identification du couple haut-parleur/canal acoustique :

1. un premier étage non linéaire constitué d'une partie linéaire simple plus une partie harmonique d'ordres 2 et 3,

$$x_{nl}(n) = x(n) + \sum_{i=0}^{L-1} \sum_{j=i}^{L-1} h_2(i, j) x(n-i) x(n-j) + \sum_{i=0}^{L-1} \sum_{j=i}^{L-1} \sum_{k=j}^{L-1} h_3(i, j, k) x(n-i) x(n-j) x(n-k) \quad (1)$$

où L représente la mémoire finie du haut-parleur. On notera la composante d'ordre 2 (resp. d'ordre 3) $x_2(n)$ (resp. $x_3(n)$).

2. un second étage linéaire de taille N identifiant la partie linéaire globale constituée de la réponse linéaire du haut-parleur convoluée avec le canal acoustique supposé linéaire,

$$\hat{d}(n) = \sum_{i=0}^{N-1} x_{nl}(n-i) w_i(n) \quad (2)$$

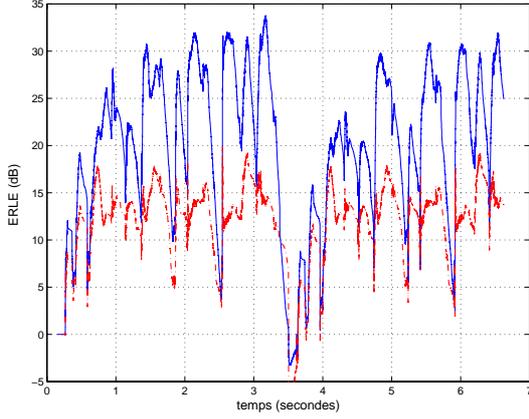


FIG. 1 – ERLE à tension maximale moins 15dB (solide) et à tension maximale (pointillés).

Ce modèle est motivé par l'impossibilité d'identifier séparément les réponses linéaires. Pour des raisons de calcul, nous utiliserons par la suite une représentation matricielle du système. On note $\mathbf{w}(n)$ le filtre adaptatif linéaire de taille $N \times 1$ et $\mathbf{x}_{nl}(n)$ son entrée de même dimension :

$$\hat{d}(n) = \mathbf{x}_{nl}^T(n) \mathbf{w}(n). \quad (3)$$

Le noyau de Volterra d'ordre 2 est représenté par le vecteur \mathbf{h}_2 de taille $L_2 \times 1$ (où $L_2 = \frac{L(L+1)}{2}$) :

$$\mathbf{h}_2^T(n) = [h_2(0,0) \ h_2(0,1) \ \dots \ h_2(L-1, L-1)]. \quad (4)$$

On associe à \mathbf{h}_2 le vecteur \mathbf{u}_2 de taille $L_2 \times 1$ tel que $x_2(n) = \mathbf{u}_2^T(n) \mathbf{h}_2(n)$:

$$\mathbf{u}_2^T(n) = [x^2(n) \ x(n)x(n-1) \ \dots \ x^2(n-L+1)]. \quad (5)$$

On définit alors la matrice \mathbf{U}_2 de dimensions $L_2 \times N$:

$$\mathbf{U}_2(n) = [\mathbf{u}_2(n) \ \mathbf{u}_2(n-1) \ \dots \ \mathbf{u}_2(n-N+1)], \quad (6)$$

telle que le vecteur $\mathbf{x}_2(n)$ s'écrive :

$$\mathbf{x}_2(n) = \mathbf{U}_2^T(n) \mathbf{h}_2(n). \quad (7)$$

En utilisant les mêmes notations, le vecteur des non-linéarités d'ordre 3 de dimension $N \times 1$ est défini par :

$$\mathbf{x}_3(n) = \mathbf{U}_3^T(n) \mathbf{h}_3(n), \quad (8)$$

où \mathbf{h}_3 est le vecteur des coefficients d'ordre 3 de dimension $L_3 \times 1$ et \mathbf{U}_3 la matrice associée de dimensions $L_3 \times N$, avec $L_3 = \frac{L(L^2+5)}{6}$.

Le vecteur d'entrée du filtre linéaire est finalement défini par l'expression matricielle :

$$\mathbf{x}_{nl}(n) = \mathbf{x}(n) + \mathbf{U}_2^T(n) \mathbf{h}_2(n) + \mathbf{U}_3^T(n) \mathbf{h}_3(n). \quad (9)$$

3 Convergence locale

L'annulation d'écho acoustique est généralement réalisée à l'aide d'un seul filtre adaptatif linéaire réactualisé par un algorithme de type gradient. Cette situation assure que le filtre de Wiener, filtre qui correspond au minimum de l'Erreur Quadratique Moyenne (EQM), sous hypothèse que le système est invariant, reste constant dans le temps :

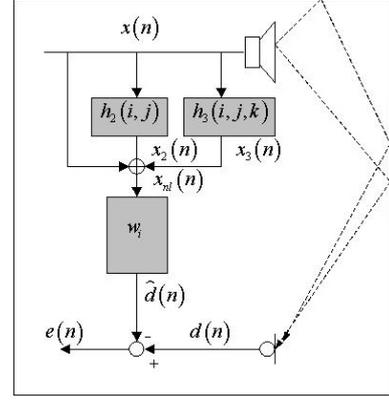


FIG. 2 – Modèle d'identification

le filtre adaptatif converge alors en moyenne, si la condition de stabilité est vérifiée, vers cette solution optimale [3]. Le cas de filtres en cascade est fondamentalement différent.

Afin d'étudier le comportement théorique des filtres de Wiener en cascade, l'ensemble haut-parleur/canal acoustique est remplacé par un système en cascade identique au modèle d'identification et de mêmes dimensions : le triplet $(\delta, \mathbf{h}_2^o, \mathbf{h}_3^o)$ en amont pour l'étage non linéaire et le filtre linéaire \mathbf{w}^o en aval (cf. figure 3). Le signal d'écho $d(n)$ est alors défini par les notations :

$$\begin{aligned} d(n) &= \mathbf{x}_{nl}^o T(n) \mathbf{w}^o \\ &= (\mathbf{x}(n) + \mathbf{x}_2^o(n) + \mathbf{x}_3^o(n))^T \mathbf{w}^o. \end{aligned} \quad (10)$$

En notant $e(n) = d(n) - \hat{d}(n)$ l'erreur en sortie du système, l'erreur quadratique moyenne J s'écrit

$$\begin{aligned} J &= E \{d^2(n)\} - 2E \{d(n) \mathbf{x}_{nl}^T(n) \mathbf{w}(n)\} \\ &\quad + E \{\mathbf{w}^T(n) \mathbf{x}_{nl}(n) \mathbf{x}_{nl}^T(n) \mathbf{w}(n)\}. \end{aligned} \quad (11)$$

Pour un filtre quelconque \mathbf{v} , le filtre optimal au sens de Wiener correspondant \mathbf{v}_{EQM} vérifie l'égalité :

$$\nabla_{\mathbf{v}} \{J(\mathbf{v})\}|_{\mathbf{v}=\mathbf{v}_{EQM}} = \mathbf{0}. \quad (12)$$

Les filtres de Wiener correspondant aux filtres \mathbf{w} et \mathbf{h}_2 , \mathbf{w}_{EQM} et $\mathbf{h}_{2,EQM}$, seront dérivés à partir du calcul du gradient de l'erreur quadratique moyenne de l'équation (12). Le cas de $\mathbf{h}_{3,EQM}$, quasi-identique à celui de $\mathbf{h}_{2,EQM}$, ne sera pas décrit. Pour des raisons de lisibilité, le temps indicé par la variable n sera omis dans les calculs dans la suite de la section 3.

3.1 Filtre linéaire \mathbf{w}

A partir de l'équation (11), le calcul du gradient $\nabla_{\mathbf{w}} \{J\}$ mène à la relation suivante :

$$\nabla_{\mathbf{w}} \{J\} = -2E \{d \mathbf{x}_{nl}\} + 2E \{\mathbf{x}_{nl} \mathbf{x}_{nl}^T\} \mathbf{w}. \quad (13)$$

En insérant l'équation (10) dans (13) et annulant ce gradient, on trouve l'expression du filtre de Wiener :

$$\mathbf{w}_{EQM} = \Gamma_{\mathbf{x}_{nl}}^{-1} \Gamma_{\mathbf{x}_{nl}, \mathbf{x}_{nl}^o} \mathbf{w}^o \quad (14)$$

où $\Gamma_{\mathbf{x}}$ (resp. $\Gamma_{\mathbf{x}, \mathbf{y}}$) est la matrice d'auto-corrélation de \mathbf{x} (resp. d'inter-corrélation des vecteurs \mathbf{x} et \mathbf{y}).

De l'équation (14) se déduisent deux remarques :

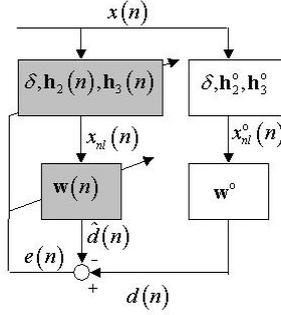


FIG. 3 – Système adaptatif (en grisé) et ensemble haut-parleur/canal acoustique modélisé (en blanc)

1. si le couple adaptatif $(\mathbf{h}_2, \mathbf{h}_3)$ du module non linéaire n'a pas convergé vers la solution réelle $(\mathbf{h}_2^o, \mathbf{h}_3^o)$, \mathbf{w} ne convergera pas (en moyenne) vers le filtre à identifier \mathbf{w}^o mais vers une version modifiée qui dépend de l'état des filtres adaptatifs \mathbf{h}_2 et \mathbf{h}_3 ,
2. dans le cas où les filtres harmoniques \mathbf{h}_2 et \mathbf{h}_3 s'adaptent continuellement, les matrices d'auto- et d'inter-corrélation $\Gamma_{\mathbf{w}_{nl}}$ et $\Gamma_{\mathbf{w}_{nl}, \mathbf{w}_{nl}^o}$ ne sont pas constantes dans le temps de même que le filtre de Wiener \mathbf{w}_{EQM} : la stabilité du filtre adaptatif $\mathbf{w}(n)$ est alors compromise.

Il convient néanmoins de modérer cette dernière observation : dans la réalité, les non-linéarités sont négligeables comparées au signal $x(n)$. Ainsi, si l'on considère en plus dans l'équation de \mathbf{w}_{EQM} que les filtres identifiés \mathbf{h}_2 et \mathbf{h}_3 sont du même ordre de grandeur que ceux du système réel \mathbf{h}_2^o et \mathbf{h}_3^o , on aboutit à l'approximation :

$$\mathbf{w}_{EQM} \simeq \mathbf{w}^o, \text{ si } \|x_2\| \ll \|x\| \text{ et } \|x_3\| \ll \|x\|, \quad (15)$$

égalité qui est d'ailleurs vérifiée dans les faits.

3.2 Filtre de Volterra d'ordre 2, \mathbf{h}_2

Le calcul du gradient $\nabla_{\mathbf{h}_2}$ est plus fastidieux : on utilise la formule développée de \mathbf{x}_{nl} donnée par l'équation (9) et celle de \mathbf{x}_{nl}^o donnée dans (10). On utilisera également le vecteur d'erreur du filtre \mathbf{w} :

$$\mathbf{e}_{\mathbf{w}}(n) = \mathbf{w}^o - \mathbf{w}(n). \quad (16)$$

On montre alors que le filtre de Wiener correspondant à \mathbf{h}_2 peut s'écrire :

$$\mathbf{h}_{2,EQM} = \mathbf{h}_2^o + \begin{aligned} & \Gamma_{\mathbf{U}_2 \mathbf{w}}^{-1} [E \{ \mathbf{U}_2 \mathbf{w} \mathbf{e}_{\mathbf{w}}^T \mathbf{x} \} \\ & + E \{ \mathbf{U}_2 \mathbf{w} \mathbf{e}_{\mathbf{w}}^T (\mathbf{x}_2^o + \mathbf{x}_3^o) \} \\ & + E \{ \mathbf{U}_2 \mathbf{w} \mathbf{w}^T (\mathbf{x}_2^o - \mathbf{x}_3) \}]. \end{aligned} \quad (17)$$

L'équation (17) est relativement complexe. Cependant, dans le cas réel, les non-linéarités représentées par les vecteurs \mathbf{x}_2^o et \mathbf{x}_3^o restent négligeables comparées au signal d'entrée \mathbf{x} (même si leurs influences sur les performances ne sont pas négligeables). De plus, si l'on considère que le vecteur d'erreur $\mathbf{e}_{\mathbf{w}}$ est du même ordre de grandeur que \mathbf{w} , les termes des lignes 2 et 3 sont négligeables comparés au terme $E \{ \mathbf{U}_2 \mathbf{w} \mathbf{e}_{\mathbf{w}}^T \mathbf{x} \}$. Il convient de noter que ces approximations ne sont valables que pour un signal coloré comme la parole car dans le cas d'un bruit blanc gaussien

centré, les espérances des puissances impaires sont nulles. On aboutit finalement à l'approximation suivante :

$$\mathbf{h}_{2,EQM} \simeq \mathbf{h}_2^o + \Gamma_{\mathbf{U}_2 \mathbf{w}}^{-1} E \{ \mathbf{U}_2 \mathbf{w} \mathbf{x}^T \} \mathbf{e}_{\mathbf{w}}. \quad (18)$$

On remarque donc qu'à l'instant n , le filtre \mathbf{h}_2 converge en moyenne vers \mathbf{h}_2^o plus un terme qui dépend linéairement de l'erreur instantanée $\mathbf{e}_{\mathbf{w}}$. Aussi, comme pour le filtre \mathbf{w} :

1. \mathbf{h}_2 ne peut converger vers sa solution \mathbf{h}_2^o que si \mathbf{w} a déjà convergé,
2. dans le cas où les deux filtres s'adaptent simultanément, le filtre de Wiener $\mathbf{h}_{2,EQM}$ n'est pas constant au cours de l'adaptation. Il en découle un comportement chaotique.

Contrairement au filtre \mathbf{w} pour lequel l'influence des non-linéarités est très faible, les simulations montrent que le filtre de Wiener $\mathbf{h}_{2,EQM}$ dépend fortement de l'erreur $\mathbf{e}_{\mathbf{w}}$, d'autant plus que cette dernière peut varier de manière importante au cours d'une communication, les changements de R.I. du canal étant relativement marqués.

4 Annulation d'écho non linéaire

On propose ici d'utiliser le modèle de la figure 2 défini par les équations (1) et (2) pour réaliser l'annulation d'écho. L'adaptation des filtres est réalisée par l'algorithme NLMS [2].

4.1 Stratégie d'adaptation

À partir des conclusions des paragraphes 3.1 et 3.2, la stratégie adoptée est la suivante :

1. afin de conserver la capacité du système à suivre les changements du canal acoustique, l'adaptation du filtre linéaire \mathbf{w} reste libre,
2. l'adaptation des filtres non linéaires \mathbf{h}_2 et \mathbf{h}_3 est gelée pendant les phases d'adaptation de \mathbf{w} , de manière à minimiser l'influence de $\mathbf{e}_{\mathbf{w}}$ et à garantir la stabilité du système.

La détection des phases d'adaptation du filtre linéaire est réalisée par seuillage de sa variance : en effet, celle-ci est minimale lorsque le filtre a convergé. Si l'on note $\overline{\mathbf{w}}(n)$ la moyenne temporelle du filtre $\mathbf{w}(n)$, l'écart-type $\sigma_{\mathbf{w}}(n)$ est estimé par un filtrage récursif du premier ordre :

$$\sigma_{\mathbf{w}}(n+1) = \begin{cases} \beta_{up} \sigma_{\mathbf{w}}(n) + (1 - \beta_{up}) \|\overline{\mathbf{w}}(n) - \mathbf{w}(n)\| & \text{si } \|\overline{\mathbf{w}}(n) - \mathbf{w}(n)\| \geq \sigma_{\mathbf{w}}(n) \\ \beta_{dn} \sigma_{\mathbf{w}}(n) + (1 - \beta_{dn}) \|\overline{\mathbf{w}}(n) - \mathbf{w}(n)\| & \text{sinon} \end{cases} \quad (19)$$

où β_{up} et β_{dn} sont des constantes proches de 1, avec la relation d'ordre $\beta_{up} \leq \beta_{dn}$.

Le filtre moyen est également estimé par filtrage récursif du premier ordre :

$$\overline{\mathbf{w}}(n+1) = \alpha \overline{\mathbf{w}}(n) + (1 - \alpha) \mathbf{w}(n) \quad (20)$$

où α est une constante proche de 1.

L'inégalité entre les coefficients d'oubli assure (i) qu'avec β_{up} correspondant à un effet mémoire faible, l'écart-type

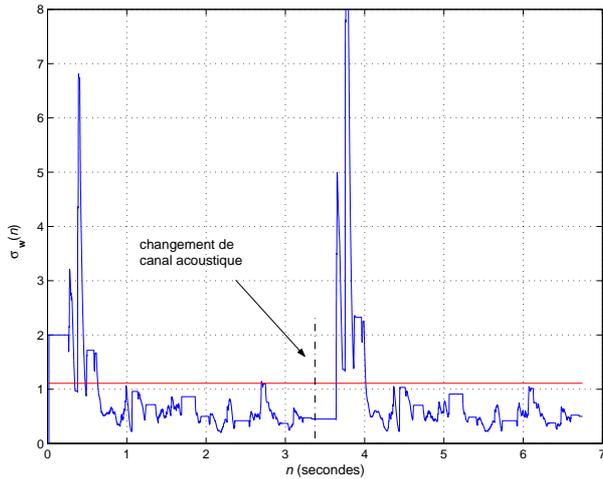


FIG. 4 – Comportement de l'écart-type $\sigma_{\mathbf{w}}$ en fonction du temps.

reste réactif à tout changement de réponse impulsionnelle du canal acoustique, (ii) qu'en prenant β_{dn} très proche de 1, l'écart-type ne décroît pas trop vite afin de geler l'adaptation des filtres non linéaires tant que le filtre \mathbf{w} n'a pas « suffisamment » convergé.

4.2 Résultats

L'algorithme développé est validé sur un signal de parole consistant en une même phrase répétée deux fois, appliquée à la tension maximale du haut-parleur avec un changement dans le canal acoustique : une avance de 4 coefficients est artificiellement introduite au début de la seconde phrase. Les tailles des filtres sont : $N = 128$, et $L = 10$, ce qui correspond à des filtres non linéaires de tailles respectives $L_2 = 55$ et $L_3 = 175$.

Les pas d'adaptation des filtres non linéaires valent 0.05. Cette valeur n'est pas critique, mais elle ne doit cependant pas être trop élevée. En effet, l'approximation de l'équation (15) n'est valable que si les termes non linéaires restent négligeables comparés au terme linéaire : une valeur élevée (tout en respectant la condition de convergence) crée des oscillations importantes des filtres \mathbf{h}_2 et \mathbf{h}_3 qui ne garantissent plus cette approximation, ce qui mène à la désadaptation de \mathbf{w} . Du fait de la faiblesse de cette valeur et également du mauvais conditionnement des matrices d'auto-corrélation des entrées des filtres \mathbf{h}_2 et \mathbf{h}_3 , la vitesse de convergence est lente. Néanmoins, leur tâche est d'identifier les paramètres d'un système non linéaire qui varie peu dans le temps : la vitesse de convergence n'est pas primordiale.

La figure 4 montre le comportement de l'écart-type $\sigma_{\mathbf{w}}$: les 2 périodes correspondant aux phases d'adaptation de \mathbf{w} (début et milieu) sont correctement détectées et les filtres non linéaires sont gelés pendant ces deux phases.

Les performances de l'algorithme proposé sont décrites figure 5. On remarque notamment que sans contrôle de l'adaptation des filtres non linéaires (courbe (c)), le système ne remplit pas son rôle d'annuleur d'écho : les filtres \mathbf{w} , \mathbf{h}_2 et \mathbf{h}_3 oscillent en permanence et sont en perpétuelle phase d'adaptation-désadaptation. En utilisant l'informa-

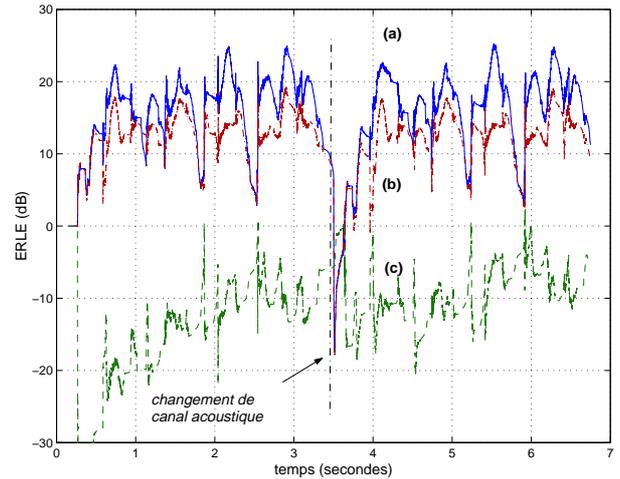


FIG. 5 – Résultats en termes d'ERLE des algorithmes (a) non linéaire avec contrôle d'adaptation de \mathbf{h}_2 et \mathbf{h}_3 (continu), (b) linéaire (trait-point) et (c) non linéaire sans contrôle d'adaptation (discontinu).

tion sur l'état de \mathbf{w} (courbe (a)), le système est stabilisé et les performances en termes d'ERLE sont supérieures à celles de l'algorithme linéaire seul (courbe (b)) : le gain moyen est de l'ordre de 4 à 5 dB.

5 Conclusion

Dans cet article est décrit un algorithme d'annulation d'écho non linéaire basé sur la modélisation du couple haut-parleur/canal acoustique par un système en cascade filtre de Volterra/filtre linéaire. Du fait de la structure cascade, les formules des filtres de Wiener associées aux différents filtres adaptatifs, dérivées dans ce papier, montrent l'existence du phénomène de convergence locale qui rend hautement instable l'adaptation conjointe des filtres adaptatifs. Une stratégie d'adaptation, basée sur la connaissance de l'état du filtre linéaire aval, a permis de stabiliser l'algorithme tout en conservant ses capacités à suivre les changements du canal acoustique : la réduction d'écho en termes d'ERLE s'avère supérieure de 4 à 5dB en comparaison de celle réalisée par l'algorithme linéaire NLMS.

Néanmoins, la charge de calcul reste pour le moment relativement importante et des travaux d'optimisation sont en cours.

Références

- [1] A. Fermo, A. Carini, G.L. Sicuranza, *Simplified Volterra filters for acoustic echo cancellation in GSM receivers*, Proc. of European Signal Processing Conference, Tampere, 2000.
- [2] A. Stenger, W. Kellermann, R. Rabenstein, *Adaptation of acoustic echo cancellers incorporating a memoryless nonlinearity*, Proc. of the International workshop on Acoustic Echo and Noise Control, Pocono Manor, 1999.
- [3] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, second edition, 1991.