



UN ALGORITHME D'ANNULATION D'ECHO SEREO DE TYPE LMS PRENANT EN COMPTE L'INTER-CORRELATION DES ENTREES

Fabrice AMAND ^①, Jacob BENESTY ^②, André GILLOIRE ^①, Yves GRENIER ^②

^① CNET LAA/TSS/CMC 2 avenue Pierre Marzin 22307 Lannion Cedex FRANCE
e_mail: amand@lannion.cnet.fr

^② TELECOM Paris 46 rue Barrault, 75634 Paris Cedex FRANCE

Le problème de l'annulation d'écho acoustique pour la téléconférence stéréophonique est rendu difficile par la corrélation existant entre les deux signaux provenant de la salle distante. Nous proposons un nouvel algorithme, de complexité proche de celle du gradient stochastique, prenant partiellement en compte la corrélation entre ces deux signaux. Cet algorithme apporte une amélioration significative par rapport au gradient stochastique bi-voie standard.

1. INTRODUCTION

Il est prévu de réaliser la prise et la restitution du son en configuration stéréophonique (et plus généralement multi-voie) dans les futurs systèmes de téléconférence, afin d'améliorer la qualité sonore de ces systèmes en permettant notamment la localisation des locuteurs distants. Par rapport au cas classique mono-voie, l'intercorrélation existant entre les signaux provenant de la salle distante dégrade la convergence des algorithmes de type gradient stochastique (LMS) mono-voie appliqués de façon directe au cas bi-voie [1]. Par conséquent, une autre forme d'algorithme doit être utilisée. L'objet de cette communication est de décrire un nouvel algorithme prenant en compte cette intercorrélation et de montrer l'amélioration correspondante des performances.

Conceptuellement, l'annulation d'écho stéréophonique peut être vue comme une généralisation de l'annulation d'écho acoustique mono-voie étendue au cas bi-voie. Un schéma simple d'annulation d'écho stéréophonique bi-voie est représenté sur la figure 1:

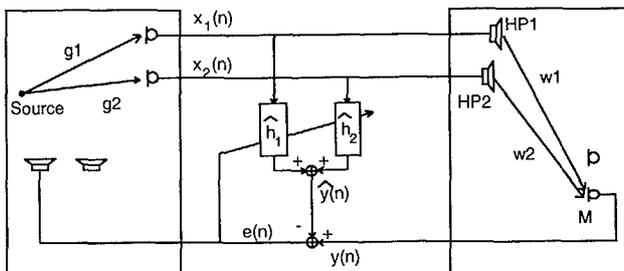


Figure 1: Schéma classique de l'annulation d'écho bi-voies

Un seul microphone est représenté car le même système est appliqué au second microphone, ainsi qu'aux microphones de la salle d'émission du son.

Acoustic echo cancellation for stereophonic teleconferencing is a difficult problem, due to the correlation between the two signals coming from the distant room. We propose a new algorithm, with complexity close to the stochastic gradient's one, which takes into account low-order correlation coefficients between these two signals. This algorithm yields significant improvements in performance, with respect to the standard two-channel stochastic gradient algorithm.

La problématique de l'annulation d'écho stéréophonique (déjà présentée dans [2]) consiste en l'identification d'un système multi-entrée formé de filtres linéaires inconnus qui dans notre cas sont représentés par les chemins acoustiques w_1 et w_2 reliant les haut parleurs (HP1,HP2) au microphone M. L'annuleur d'écho stéréophonique essaie de modéliser le système inconnu par une paire de filtres adaptatifs (h_1, h_2).

Il apparaît clairement sur la figure 1 que les signaux x_1 et x_2 sont issus de la même source et que, par conséquent, ils sont corrélés, ce qui conduit à un problème mal conditionné [1]. En pratique, grâce au bruit ambiant et au fait que la longueur des réponses impulsionnelles g_1 et g_2 n'est pas liée à la longueur des filtres de modélisation h_1 et h_2 , la matrice de covariance des signaux x_1 et x_2 est toujours inversible. Cette inversibilité permet après convergence l'identification séparée des canaux w_1 et w_2 par les filtres h_1 et h_2 , par la minimisation d'un critère quadratique quelconque. Ainsi, un algorithme comme le MCRR (moindres carrés récursifs rapides), dont les caractéristiques sont indépendantes du conditionnement de la matrice de covariance, a un comportement similaire au MCRR mono-voie [3]. Mais cet algorithme, avec une complexité de $28L$ (L : taille des filtres) dans le cas bi-voie, est inacceptable pour l'application de téléconférence. Aussi, nous nous intéressons aux algorithmes de type gradient stochastique, qui sont de complexité nettement inférieure mais sont sensibles aux effets du conditionnement matriciel.

2. EFFET DU CONDITIONNEMENT

Nous avons étudié le conditionnement des matrices de covariance des signaux d'entrée correspondant aux cas mono et bi-voie, dans le cas de la parole et du bruit USASI (à spectre moyen de parole). Avec du bruit USASI en entrée, les performances du NLMS bi-voie sont énormément dégradées par rapport au NLMS mono-voie. Dans le cas de la parole, la matrice de corrélation mono-voie étant déjà mal conditionnée,



les termes d'inter-corrélation ont moins d'influence que dans le cas du bruit USASI. Ceci explique que les algorithmes mono et bi-voies ont des performances plus proches dans le cas de la parole que dans le cas du bruit USASI [4]. Cette différence provient du fait que le LMS bi-voie ne prend pas en compte l'inter-corrélation des entrées, c'est une solution sous-optimale. Aussi, nous en déduisons que des algorithmes de type stochastique prenant en compte la corrélation entre les signaux d'entrée doivent résoudre le problème d'une façon plus proche de l'optimalité quel que soit le type de signal source. Nous proposons un nouvel algorithme tenant compte des premiers coefficients d'inter-corrélation, que nous appelons algorithme LMS-étendu (ELMS), et qui peut être considéré comme une approximation très simple du MCR à deux voies.

3. L'ALGORITHME LMS-ETENDU

Pour obtenir l'algorithme LMS-étendu, nous allons nous inspirer de la méthode mono-voie d'obtention du LMS à partir du MCR, c'est à dire par approximation de la matrice de corrélation. A partir de l'équation classique d'adaptation des filtres des moindres carrés bi-voie [5]:

$$e(n) = y(n) - \begin{pmatrix} H_1^t(n) & H_2^t(n) \end{pmatrix} \begin{pmatrix} X_1(n) \\ X_2(n) \end{pmatrix}$$

$$\begin{pmatrix} H1(n+1) \\ H2(n+1) \end{pmatrix} = \begin{pmatrix} H1(n) \\ H2(n) \end{pmatrix} + \begin{pmatrix} R_{11}(n) & R_{12}(n) \\ R_{21}(n) & R_{22}(n) \end{pmatrix}^{-1} \begin{pmatrix} X1(n) \\ X2(n) \end{pmatrix} \cdot e(n)$$

où $R_{ij}(n)$, $i,j=(1,2)$, est une estimation de la matrice de covariance de dimension $L \times L$ entre les signaux x_i et x_j , correspondant au critère minimisé par l'algorithme.

Nous allons écrire explicitement l'inverse de la matrice de covariance bi-voie (pour simplifier les notations nous posons $R_{ij}(n)=R_{ij}$):

$$\begin{pmatrix} H1(n+1) \\ H2(n+1) \end{pmatrix} = \begin{pmatrix} H1(n) \\ H2(n) \end{pmatrix} + \begin{pmatrix} (R_{11}-R_{12} \cdot R_{22}^{-1} \cdot R_{21})^{-1} & \underline{0}_L \\ \underline{0}_L & (R_{22}-R_{21} \cdot R_{11}^{-1} \cdot R_{12})^{-1} \end{pmatrix} \begin{pmatrix} I_L & -R_{12} \cdot R_{22}^{-1} \\ -R_{21} \cdot R_{11}^{-1} & I_L \end{pmatrix} \begin{pmatrix} X1(n) \\ X2(n) \end{pmatrix} \cdot e(n)$$

Avec $\underline{0}_L$ la matrice zéro de dimension $L \times L$ et I_L la matrice identité de dimension $L \times L$. En approchant les matrices d'auto-corrélation et d'inter-corrélation des signaux d'entrée par les estimations des coefficients de corrélation d'ordre 0:

$$R_{11}^{-1} = \frac{1}{\rho_{11}} \cdot I_L \quad R_{12} = \rho_{12} \cdot I_L$$

$$R_{22}^{-1} = \frac{1}{\rho_{22}} \cdot I_L \quad R_{21} = \rho_{21} \cdot I_L$$

Nous obtenons le LMS-étendu:

$$e(n) = y(n) - \begin{pmatrix} H_1^t(n) & H_2^t(n) \end{pmatrix} \begin{pmatrix} X_1(n) \\ X_2(n) \end{pmatrix}$$

$$H1(n+1) = H1(n) + pas1 \cdot (X1(n) - ro \cdot \frac{\rho_{12}}{\rho_{22}} \cdot X2(n)) \cdot e(n)$$

$$H2(n+1) = H2(n) + pas2 \cdot (X2(n) - ro \cdot \frac{\rho_{21}}{\rho_{11}} \cdot X1(n)) \cdot e(n)$$

Avec comme pas d'adaptation:

$$pas1 = \mu_1 \cdot (Z1^t(n) \cdot Z1(n))^{-1}$$

$$Z1(n) = X1(n) - ro \cdot \frac{\rho_{12}}{\rho_{22}} \cdot X2(n)$$

$$pas2 = \mu_2 \cdot (Z2^t(n) \cdot Z2(n))^{-1}$$

$$Z2(n) = X2(n) - ro \cdot \frac{\rho_{21}}{\rho_{11}} \cdot X1(n)$$

Où μ_1, μ_2 et ro sont respectivement les facteurs de pondération des pas d'adaptation et des coefficients d'inter-corrélation.

Pour déterminer les conditions de stabilité de l'algorithme proposé, nous utilisons la condition nécessaire de convergence:

$$E[\varepsilon(n)] \leq E[\varepsilon(n)]$$

où $\varepsilon(n)$ est l'erreur de filtrage a posteriori:

$$\varepsilon(n) = y(n) - \begin{pmatrix} H_1^t(n+1) & H_2^t(n+1) \end{pmatrix} \begin{pmatrix} X_1(n) \\ X_2(n) \end{pmatrix}$$

Cette condition exige que l'erreur a posteriori soit inférieure à l'erreur a priori en moyenne. Après développement, nous arrivons aux conditions de convergence:

$$0 < \mu_1 + \mu_2 < 2 \quad \text{pas d'adaptation des coefficients}$$

$$0 \leq ro < 1 \quad \text{pondération des coefficients d'inter-corrélation}$$

On peut remarquer que, pour le cas $ro=0$ le LMS-étendu correspond au LMS appliqué directement au cas bi-voies.

Cet algorithme a une complexité de $6L$ multiplications et $6L$ additions (à comparer au $4L$ du LMS) avec une mise à jour des coefficients d'auto-corrélation, d'inter-corrélation et des pas d'adaptation par lissage exponentiel remplaçant les produits scalaires:

$$\rho_{ij}(n) = w \cdot \rho_{ij}(n-1) + (1-w) \cdot x_i(n) \cdot x_j(n)$$

$$0 < w < 1, w \text{ étant le facteur d'oubli}$$

Ici, nous sommes en présence d'un algorithme qui décorrèle explicitement par des coefficients les deux vecteurs X_1 et X_2 . Ce n'est pas un algorithme de projection "classique" comme nous le connaissons dans le cas mono-voie. L'algorithme LMS-étendu ne fait pas une projection dans le temps d'un vecteur concaténé (X_1, X_2) à l'instant n et $n-1$, qui ne prendrait pas en compte l'inter-corrélation des deux entrées (qui est le problème spécifique du bi-voie). Le LMS-étendu fait la mise à jour des filtres en utilisant la composante décorrélée de chaque vecteur. En effet, nous pouvons interpréter géométriquement:

$$X1(n) - ro \cdot \frac{\rho_{12}}{\rho_{22}} \cdot X2(n)$$

comme étant la projection de X_1 (via le facteur ro) sur la composante orthogonale à X_2 par rapport à X_1 . Notons que le comportement de l'algorithme est lié à l'erreur qui est elle-même dépendante de X_1 et X_2 .

4. COMPORTEMENT DU LMS-ETENDU

Nous avons simulé l'algorithme LMS-étendu dans le cas du bruit USASI et de la parole. Les simulations sont "académiques": les réponses de la salle sont de 256 points, les filtres adaptatifs sont aussi de 256 points. Un bruit blanc est ajouté en sortie (nous avons un rapport signal sur bruit de 60 dB). La prise de son est effectuée dans la salle distante au moyen de deux microphones proches l'un de l'autre, ce qui conduit à une corrélation importante de X1 et X2. L'algorithme LMS-étendu est donc supposé pouvoir apporter une amélioration dans ce cas. Sur les figures 2 et 3 nous avons comparé les algorithmes LMS et LMS-étendu. Les deux algorithmes sont réglés pour avoir la vitesse de convergence la plus rapide possible.

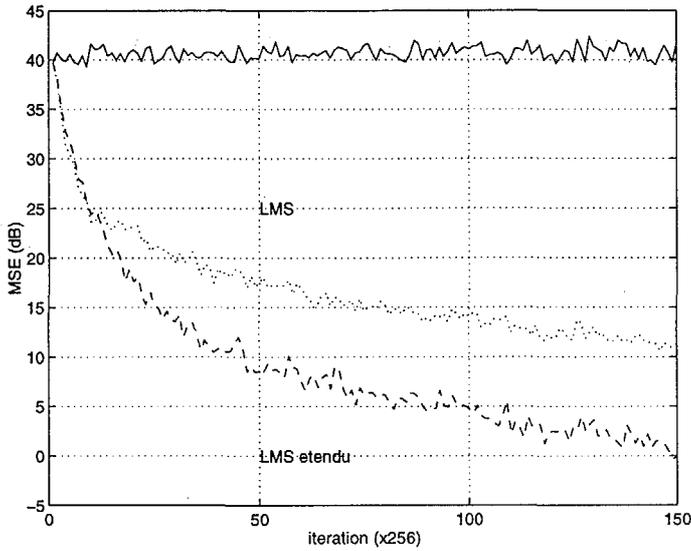


Figure2: Comparaison LMS-ELMS avec du bruit USASI.

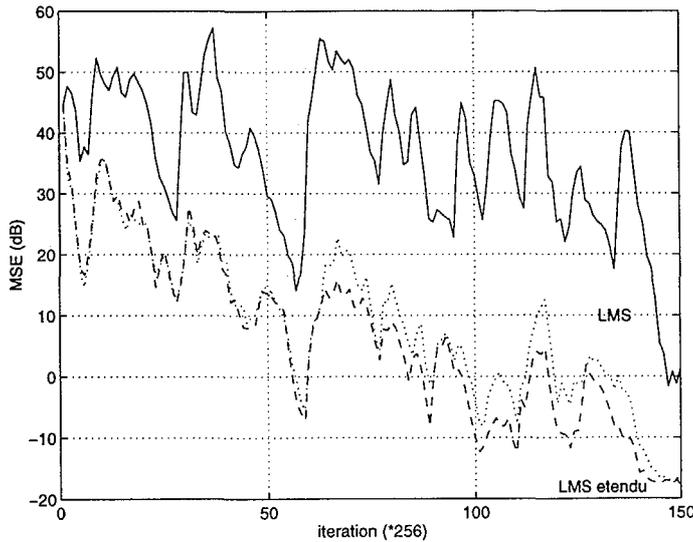


Figure3: Comparaison LMS-ELMS avec de la parole.

Pour ces deux simulations, nous avons pris $\mu = 0.5$ pour le LMS et $\mu_1 = \mu_2 = 0.5$, $ro = 0.9$ pour le ELMS.

Nous pouvons constater que l'algorithme LMS-étendu a un meilleur comportement que le LMS bi-voie. Cette propriété a

tendance à s'atténuer lorsqu'on augmente la taille des filtres, ainsi que dans certaines configurations de prise de son, où un seul coefficient n'est alors pas suffisant pour décorrélérer les sources. Or, c'est précisément le cas des réponses longues qui nous intéresse pour la téléconférence ! Il faut donc voir cet algorithme, non pas comme une fin en soit, mais comme un premier pas vers une solution optimale prenant en compte le problème spécifique de l'annulation d'écho multi-voie: l'inter-corrélation des signaux d'entrée.

A partir de cet algorithme, qui prend en compte le problème intrinsèque du multi-voie, les techniques mono-voie classiques peuvent être appliquées, comme par exemple le passage dans le domaine fréquentiel (dans ce cas nous parlons de fonction de cohérence), ou bien l'extension à un ordre supérieur des coefficients d'inter-corrélation, c'est à dire non plus des coefficients, mais des matrices. Nous allons donner dans ce dernier cas un algorithme LMS étendu à l'ordre 2.

5. EXTENSION A L'ORDRE 2

Une fois le problème du multi-voie pris en compte sans aucune hypothèse sur les signaux de la prise de son, nous pouvons généraliser notre algorithme. Nous proposons une version intuitive à l'ordre 2 (sans justification de cette solution par un critère mathématique).

Algorithme LMS-étendu à l'ordre 2:

$$e(n) = y(n) - \begin{pmatrix} H_1^t(n) & H_2^t(n) \end{pmatrix} \cdot \begin{pmatrix} X_1(n) \\ X_2(n) \end{pmatrix}$$

$$H1(n+1) = H1(n) +$$

$$pas1. (\underline{X}1(n) - ro. \underline{X}2(n)) \cdot \begin{bmatrix} \rho22(n) & \rho22(n-1) \\ \rho22(n-1) & \rho22(n) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \rho21(n) & \rho21(n-1) \\ \rho21(n+1) & \rho21(n) \end{bmatrix} \cdot \underline{e}(n)$$

$$H2(n+1) = H2(n) +$$

$$pas2. (\underline{X}2(n) - ro. \underline{X}1(n)) \cdot \begin{bmatrix} \rho11(n) & \rho11(n-1) \\ \rho11(n-1) & \rho11(n) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \rho12(n) & \rho12(n-1) \\ \rho12(n+1) & \rho12(n) \end{bmatrix} \cdot \underline{e}(n)$$

Avec:

$$\underline{X}i(n) = [Xi(n) \quad Xi(n-1)]$$

$$\rho_{ij}(n-p+k) = Xi^t(n-k) \cdot Xj(n-p)$$

$$\underline{e}(n) = \begin{bmatrix} e(n) \\ e(n-1) \end{bmatrix}$$

pour les pas de normalisation, nous avons:

$$pas1 = \mu_1 \cdot (Z1^t(n) \cdot Z1(n))^{-1}$$

$$Z1(n) = (\underline{X}1(n) - ro. \underline{X}2(n)) \cdot \begin{bmatrix} \rho22(n) & \rho22(n-1) \\ \rho22(n-1) & \rho22(n) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \rho21(n) & \rho21(n-1) \\ \rho21(n+1) & \rho21(n) \end{bmatrix} \cdot \underline{e}(n)$$

$$pas2 = \mu_2 \cdot (Z2^t(n) \cdot Z2(n))^{-1}$$

$$Z2(n) = (\underline{X}2(n) - ro. \underline{X}1(n)) \cdot \begin{bmatrix} \rho11(n) & \rho11(n-1) \\ \rho11(n-1) & \rho11(n) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \rho12(n) & \rho12(n-1) \\ \rho12(n+1) & \rho12(n) \end{bmatrix} \cdot \underline{e}(n)$$

Pour constater l'amélioration due au passage d'un coefficient à une matrice, nous allons comparer l'algorithme LMS-étendu à l'ordre 1 et à l'ordre 2, dans les mêmes conditions qu'au paragraphe 4 (figure 4 et 5).



Pour ces deux simulations, nous avons pris $\mu_1 = \mu_2 = 0.5$, $ro = 0.9$ pour le ELMS à l'ordre 1 et $\mu_1 = \mu_2 = 0.1$, $ro = 1$ pour le ELMS à l'ordre 2.

Nous pouvons constater que la prise en compte de l'inter-corrélation entre les signaux d'une façon plus complète apporte une amélioration. La complexité de cet algorithme est de l'ordre de 8L multiplications et additions, les coefficients d'auto et d'inter-corrélation ainsi que les pas d'adaptation étant mis à jour par lissage exponentiel.

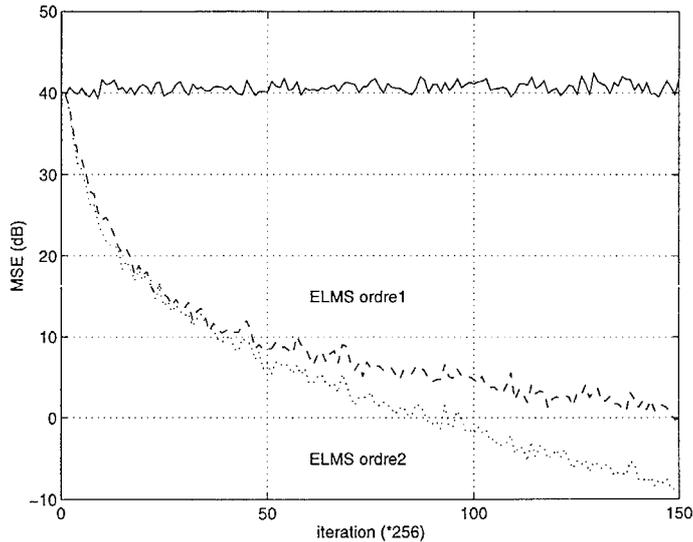


Figure 4: Comparaison ELMS avec du bruit USASI.

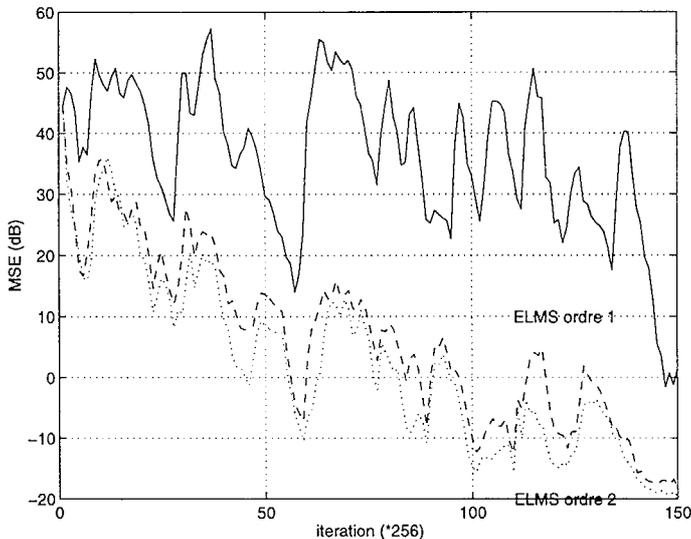


Figure 5: Comparaison ELMS avec de la parole.

6. CONCLUSION

Nous avons décrit le principal problème du passage de l'annulation d'écho acoustique mono-voie au cas bi-voie dans le cas des algorithmes à gradient stochastique. Une façon de résoudre le problème est de prendre en compte l'inter-corrélation des signaux d'entrée. Nous proposons dans ce but un algorithme -le LMS-étendu- qui constitue une première étape simple vers la détermination d'une meilleure solution, qui peut être obtenue par exemple par un passage dans le domaine fréquentiel, ou par extension de coefficients en matrices.

L'extension de cet algorithme à un ordre plus élevé conduit à une complexité de calcul croissante; il semble néanmoins possible d'obtenir une version rapide au moyen de techniques analogues à celle utilisées pour obtenir les algorithmes de projection rapides.

Références

- [1] Y.Mahieux, A. Gilloire, F. Khalil, "Annulation d'Echo en Téléconférence Stéréophonique", 14ème Grets, Sept. 1993, Juan les Pins, France.
- [2] M.Sondhi and D.R.Morgan, "Acoustic Echo Cancellation for Stereophonic Teleconferencing", Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, NY 1991, USA.
- [3] J.Benesty, F.Amand, A.Gilloire, Y.Grenier, "Adaptive filtering algorithms for stereophonic acoustic echo cancellation", ICASSP, May 1995, Detroit, USA.
- [4] F.Amand, J.Benesty, A.Gilloire, Y.Grenier, "Multi-channel Acoustic Echo Cancellation", 4th International Workshop on Acoustic Echo and Noise Control, Roros, June 1993, Norway.
- [5] M.Bellanger "Analyse des signaux et filtrage numérique adaptatif", p260, ed. Masson 1989, France.