



ANNULATION D'ECHO EN TELECONFERENCE STEREOPHONIQUE

Yannick MAHIEUX, André GILLOIRE, Fadi KHALIL

Centre National d'Etudes des Télécommunications

Route de Trégastel - 22300 LANNION

RÉSUMÉ

En annulation d'écho stéréophonique, la corrélation entre les deux signaux alimentant les haut parleurs pose des problèmes nouveaux par rapport au cas classique mono voie. Cet article présente et compare quelques techniques pour résoudre ces problèmes. Bien qu'aucune d'entre elles ne donne pour l'instant de résultats satisfaisants, elles indiquent des pistes intéressantes. Ces techniques sont : l'extension directe de l'annuleur mono voie (structure dite conventionnelle) avec adaptation par le gradient stochastique temporel ou fréquentiel (GMDF) et une structure avec décorrélation préalable des deux signaux d'entrée. La structure conventionnelle avec le GMDF donne les meilleurs résultats grâce à la convergence rapide de cet algorithme.

I INTRODUCTION

Les systèmes de téléconférence actuels transmettent la parole sur une seule voie audio. L'annulation d'écho acoustique est alors réalisée classiquement au moyen d'un filtre adaptatif placé entre l'entrée (haut parleur) et la sortie (microphone) de la salle. La transmission sur une seule voie ne permet pas de reconstituer à l'extrémité distante, l'ambiance sonore en particulier la localisation des locuteurs. La transmission d'une information plus complète, par exemple deux voies comme en stéréophonie, devrait rendre possible la localisation et un meilleur rendu de l'environnement sonore.

Dans ce cas, l'annulation d'écho devient plus complexe puisqu'il s'agit d'identifier la fonction de transfert d'un système multi entrées / multi sorties. Cette identification est rendue délicate par la corrélation existant entre les signaux d'entrée, c'est-à-dire les signaux captés par les microphones dans la salle distante. Des expérimentations sur une structure découlant de l'extension directe du principe de l'annulation d'écho mono voie au cas stéréo, ont par exemple montré que les filtres identifiés par les algorithmes adaptatifs ne correspondent pas en général

ABSTRACT

In stereophonic echo cancellation, the cross correlation between the two loudspeaker signals is critical for the convergence of the algorithms. This paper proposes and compares some techniques to solve this problem; however, those techniques are not yet fully satisfactory. In particular, we describe a conventional structure directly derived from the mono channel echo canceller and which uses either time or frequency domain LMS adaptive filters. Another structure is based on a preliminary decorrelation of the two input signals. It is shown that the conventional structure with the frequency domain LMS (GMDF) yields the best performance, due to the faster convergence of the GMDF.

aux réponses impulsionnelles réelles des chemins d'écho[1].

Ce problème a suscité quelques autres études. Notons d'abord les travaux effectués aux Bell Labs [2] sur la structure conventionnelle. Cette structure est également analysée par A. Hirano & A. Sugiyama [3] qui proposent une alternative nécessitant un seul filtre adaptatif par microphone. Les deux filtres utilisent celui des deux signaux qui est en avance par rapport à l'autre. Ce signal est sélectionné à l'aide d'un détecteur opérant sur l'inter corrélation des signaux d'entrée. Cette structure présente l'avantage d'être compacte et assure une convergence rapide. Toutefois, ses performances restent dépendantes de l'environnement acoustique dans la salle distante et en particulier, sont fortement dégradées lors d'un changement de locuteur.

Le présent article est limité au cas de l'annulation d'écho stéréophonique. Dans un premier temps la structure conventionnelle analogue à celle proposée dans [1] et [2] est rappelée et analysée (section II). Une solution procédant par décorrélation préalable du signal sur chaque voie est présentée en section III. Enfin on analyse les performances d'une structure conventionnelle utilisant l'algorithme d'adaptation fréquentielle.



II METHODE CONVENTIONNELLE D'ANNULATION D'ECHO STEREOFONIQUE

La figure 1 représente un système généralisant de façon conventionnelle l'annulation d'écho mono voie au cas stéréo. Pour conserver des notations et des figures simples, on ne considère qu'un seul signal microphonique $y(n)$. L'analyse suivante est applicable au signal provenant de l'autre microphone. Le signal $y(n)$ peut être modélisé comme étant la somme des filtrages des deux signaux haut parleurs $x_1(n)$ et $x_2(n)$ par les réponses impulsionnelles des deux canaux acoustiques "haut parleurs microphone 1", $h_1(n)$ et $h_2(n)$.

Soient $H_1(n)$ et $H_2(n)$ les deux vecteurs colonnes de dimension L , composés des L premiers échantillons des réponses impulsionnelles de la salle. Soient $X_1(n)$ et $X_2(n)$ deux vecteurs colonnes de dimension L composés des L derniers échantillons des signaux x_1 et x_2 . En supposant ces réponses impulsionnelles invariantes dans le temps, le signal microphonique s'écrit:

$$y(n) = H'_1 \cdot X_1(n) + H'_2 \cdot X_2(n) \quad (1)$$

où ' désigne la transposition des vecteurs.

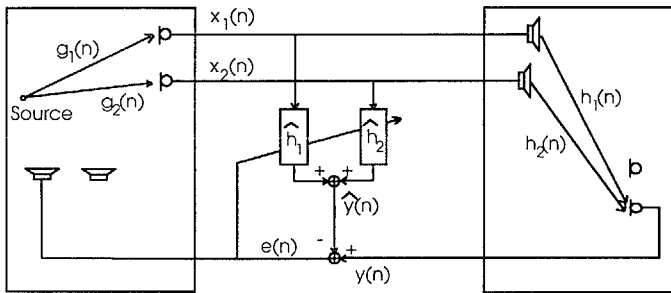


Figure 1. Méthode conventionnelle d'annulation d'écho en téléconférence stéréophonique

Dans la méthode conventionnelle, l'annuleur d'écho synthétise une estimation de $y(n)$ à l'aide de deux filtres transverses $\hat{H}_1(n)$ et $\hat{H}_2(n)$:

$$\hat{y}(n) = \hat{H}'_1(n) \cdot X_1(n) + \hat{H}'_2(n) \cdot X_2(n) \quad (2)$$

On se sert de l'erreur $e(n) = y(n) - \hat{y}(n)$ pour adapter les deux filtres $\hat{H}_1(n)$ et $\hat{H}_2(n)$ (voir figure 1). Les filtres qui minimisent l'erreur quadratique moyenne $E\{e^2(n)\}$ obéissent au système d'équations suivant :

$$\begin{aligned} \Gamma_{x_1x_1} \cdot \hat{H}_1(n) + \Gamma_{x_1x_2} \cdot \hat{H}_2(n) &= \Gamma_{yx_1} \\ \Gamma_{x_2x_1} \cdot \hat{H}_1(n) + \Gamma_{x_2x_2} \cdot \hat{H}_2(n) &= \Gamma_{yx_2} \end{aligned} \quad (3)$$

$\Gamma_{x_1x_1}$ et $\Gamma_{x_2x_2}$ sont les matrices de covariance de taille $L \times L$ des signaux x_1 et x_2 , $\Gamma_{x_1x_2}$ et $\Gamma_{x_2x_1}$ sont les matrices inter corrélation entre x_1 et x_2 , Γ_{yx_1} et Γ_{yx_2} sont les vecteurs ($L \times 1$) inter corrélation entre y et x_1 et x_2 .

D'après (3), il apparaît que si les matrices $\Gamma_{x_1x_1}$ et $\Gamma_{x_2x_1}$ ne sont pas nulles, c'est-à-dire si les signaux x_1 et x_2 sont corrélés, la solution du problème dépend des propriétés des signaux d'entrée contrairement à l'annulation d'écho mono voie. De plus dans le cas particulier où les deux signaux x_1 et x_2 sont issus du même signal x par filtrage linéaire (modèle simple de situation réelle), le système (3) dégénère en deux équations équivalentes et donc n'a pas de solution unique. Les réponses impulsionnelles identifiées dépendent alors des conditions initiales imposées aux filtres adaptatifs, et la convergence est perturbée.

On a comparé la convergence de l'annuleur d'écho mono voie avec celle de la structure de la figure 1 lorsque les filtres sont adaptés par l'algorithme NLMS. Les signaux d'entrée x_1 et x_2 échantillonnés à 16 kHz sont construits à partir d'un même signal x par filtrages linéaires g_1 et g_2 (réponses de salle de longueur 1024) pondérés de façon à ce que les puissances des deux signaux soient proches. Les filtres h_1 et h_2 sont des réponses impulsionnelles de salle tronquées ($L = 256$) correspondant aux canaux acoustiques entre deux haut-parleurs et un même microphone. En annulation d'écho mono voie, on utilise comme entrée le signal x_1 et comme sortie x_1 filtré par h_1 . La figure 2 présente la convergence des deux annuleurs d'écho; le signal x est un bruit à spectre moyen de parole. Les pas d'adaptation ont été choisis identiques.

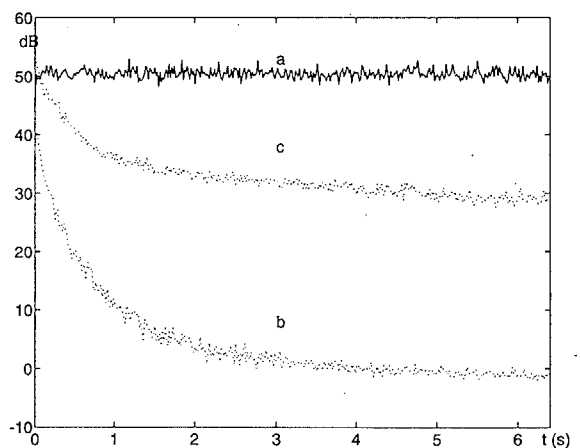


Figure 2. Annulation d'écho mono et stéréo a) Energie de l'écho, b) énergie du résidu en annulation d'écho mono voie, c) énergie du résidu en annulation d'écho stéréo, structure conventionnelle.

On observe que la convergence est fortement ralentie dans le cas stéréo et que le régime asymptotique est loin d'être atteint au bout de plus de 100.000

itérations. D'autre part, les réponses impulsionnelles des filtres adaptatifs à la fin de la simulation sont une combinaison des réponses des deux canaux acoustiques h_1 et h_2 , comme le laisse prévoir le système (3) (voir également [1]). Par ailleurs, des expériences avec du bruit blanc en entrée ont montré que l'erreur asymptotique en stéréo est de l'ordre de deux fois celle atteinte en mono voie.

III IDENTIFICATION AVEC DECORRELATION PREALABLE

D'après (3), il apparaît que si les signaux x_1 et x_2 sont décorrelés, la solution du problème est unique ($\hat{h}_1 = h_1$ et $\hat{h}_2 = h_2$). Il résulte de cette observation la structure de la figure 3 dans laquelle on construit un signal décorrelé, ε_1 , par rapport à x_1 et un signal ε_2 , décorrelé par rapport à x_2 .

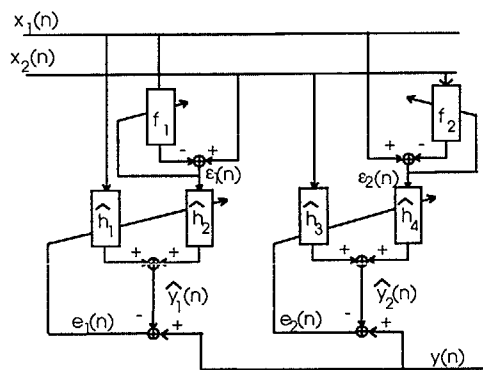


Figure 3. Annulation d'écho avec décorrélation préalable

Cette structure s'apparente aux techniques "référence bruit seul" utilisées en traitement multi capteurs [4]. Les signaux ε_1 , x_1 , ε_2 et x_2 alimentent les filtres adaptatifs \hat{h}_1 à \hat{h}_4 qui permettent l'identification des vraies réponses impulsionnelles. On peut en effet montrer que si x_1 et ε_1 sont décorrelés ainsi que x_2 et ε_2 , alors $\hat{h}_2 = h_2$ et $\hat{h}_4 = h_1$. L'estimée de l'écho, $\hat{y}(n)$, peut donc être générée par sommation des filtrages de x_1 par \hat{h}_4 et de x_2 par \hat{h}_2 . On notera que la complexité de mise en oeuvre est fortement augmentée par rapport à la structure conventionnelle, d'autant plus que les filtres décorrélateurs doivent avoir une taille théorique de $2L-1$ coefficients. D'autre part, la mise en cascade des filtres adaptatifs décorrélateurs et identificateurs conduit à des interactions qui perturbent la convergence de l'anneur d'écho.

On a étudié expérimentalement le comportement d'une structure réduite comprenant une seule branche de la figure 3 avec $e_1(n)$ pour résidu global. Il a été observé lors d'expériences avec des signaux stationnaires que l'interaction entre les filtres adaptatifs perturbe le fonctionnement de l'anneur. Même avec des pas d'adaptation très faibles, le comportement asymptotique est dégradé et la réponse impulsionnelle qui en théorie est correctement identifiée est en fait fortement bruitée.

Pour des signaux non stationnaires comme la parole et pour les mêmes réglages des filtres identificateurs, on observe une dégradation importante des performances par rapport à la structure conventionnelle.

En conclusion, la mise en oeuvre pratique de la structure décorrélatrice pose des problèmes que nous n'avons pas résolus jusqu'à présent.

IV ADAPTATION DANS LE DOMAINE FREQUENTIEL

Compte tenu du ralentissement de la convergence provoqué par la corrélation entre les signaux x_1 et x_2 , il semble intéressant d'utiliser des algorithmes à convergence rapide pour améliorer les performances de la structure conventionnelle comme suggéré dans [2]. Parmi les algorithmes possibles c'est le Generalized Multi-Delay Adaptive Filter (GMDF) [5] qui a été retenu. Cet algorithme est un NLMS fréquentiel par bloc qui présente à la fois des bonnes performances en convergence et une complexité inférieure à celle du NLMS temporel.

La structure testée est équivalente à celle de la figure 1, les deux convolutions et l'adaptation des filtres étant réalisées dans le domaine fréquentiel. On notera que la complexité est inférieure à celle de deux algorithmes GMDF indépendants puisqu'une seule erreur globale est synthétisée. Les résultats présentés ci dessous ont été obtenus avec un algorithme GMDF contraint ayant un facteur de sur échantillonnage de 4 [5]. Ces résultats sont comparés à ceux de la même structure conventionnelle avec adaptation des deux filtres par le NLMS. On a pris soin de choisir dans les deux cas, des pas d'adaptation dont le rapport à la valeur maximale assurant la convergence est identique. Ainsi les algorithmes sont comparés dans des conditions de réglage semblables.

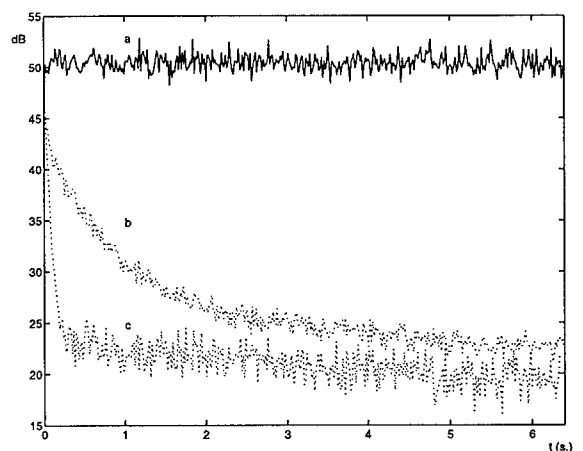


Figure 4 - Annulation d'écho - Structure conventionnelle a) Energie de l'écho, b) Energie du résidu. Adaptation des deux filtres par le NLMS, c) Adaptation par le GMDF



La figure 4 compare les deux algorithmes dans le cas de filtres h_1 et h_2 de taille 256. Les signaux x_1 et x_2 sont identiques à ceux utilisés dans l'expérience de la figure 2. On observe la convergence bien plus rapide du GMDF. Au bout de 10^5 itérations, les régimes asymptotiques semblent atteints et l'erreur asymptotique du GMDF est inférieure à celle du NLMS.

En figure 5, on compare les deux algorithmes pour un signal de parole en entrée du système. Les signaux x_1 et x_2 ont été obtenus à l'aide des mêmes filtrages g_1 et g_2 que dans l'expérience précédente. Les filtres h_1 et h_2 ont une taille de 1024 échantillons. On observe là aussi la convergence bien plus rapide du GMDF confirmée par l'écoute.

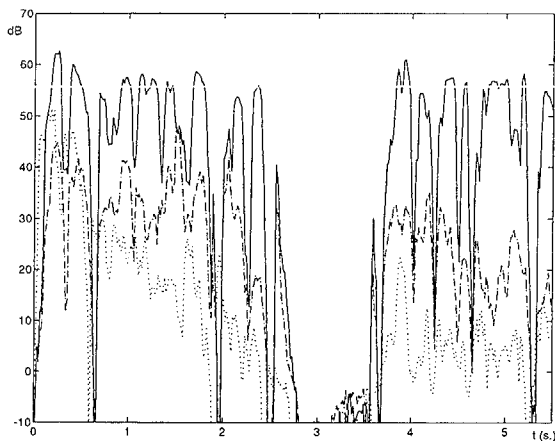


Figure 5 - Annulation d'écho - Structure conventionnelle
 Trait continu : Energie de l'écho. Tirets : Energie du résidu, Adaptation par le NLMS. Pointillés : Adaptation par le GMDF

La dépendance des solutions du système (3) par rapport à l'intercorrélacion des signaux provenant de la salle distante pose le problème de la poursuite des variations temporelles de cette corrélation dues par exemple aux mouvements du locuteur distant. Une non stationnarité a été simulée en imposant une variation linéaire du gain sur les signaux x_1 et x_2 . Le signal x_1 est multiplié par un gain variant de 1 à 2 en l'espace de 32 ms alors que le signal x_2 est pondéré par le gain inverse.

Les conditions de l'expérience sont les mêmes que celles de la figure 4 hormis le signal x à l'origine de x_1 et x_2 qui est cette fois-ci un bruit blanc (ceci de façon à atteindre la convergence plus rapidement). Pour rendre visible l'erreur de poursuite, la non stationnarité a été appliquée à partir d'un instant où les algorithmes sont proches de la convergence asymptotique (trait horizontal sur la figure 6). Le GMDF donne une erreur de poursuite moyenne inférieure de 0,5 dB à celle du NLMS.

Au vu de ces résultats, il apparaît que le GMDF améliore nettement les performances de la structure conventionnelle dans les conditions testées, mais il ne résout pas le problème de la poursuite des évolutions acoustiques dans la salle distante.

Une possibilité intéressante serait d'exploiter les caractéristiques fréquentielles des deux signaux d'entrée. Par exemple, pour chaque composante fréquentielle, l'adaptation pourrait être pilotée par la corrélation entre x_1 et x_2 , mesurée par la fonction de cohérence.

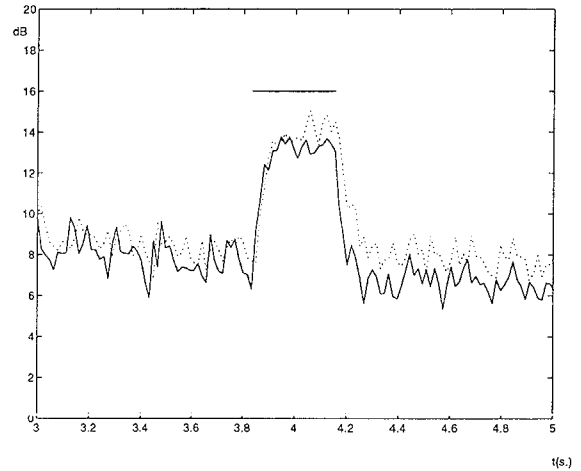


Figure 6 - Energie du résidu lors d'une non stationnarité
 Trait continu : structure conventionnelle avec adaptation par GMDF. Pointillés : Adaptation par le NLMS

V CONCLUSION

Le problème de l'annulation d'écho stéréophonique a été exposé et les difficultés liées à la corrélation entre les signaux d'entrée ont été mises en évidence lors d'expérimentations de diverses structures d'annuleurs. L'emploi d'algorithmes à convergence rapide tel le GMDF, améliore les performances de la structure conventionnelle dont l'avantage majeur par rapport à la structure décorrélatrice est la simplicité de mise en oeuvre. Toutefois une analyse théorique plus poussée du problème reste à faire pour aboutir à des algorithmes d'annulation d'écho aux performances indépendantes des variations d'acoustique dans la salle distante.

REFERENCES

- [1] : F. Khalil : "Antenne acoustique pour la prise de son dans les salles de téléconférence", thèse INPG, 1992
- [2] : M. Sondhi & D. Morgan : "Acoustic Echo Cancellation for Stereophonic Teleconferencing". IEEE Workshop on Audio and Acoustics, Mohonk Mountain House, NY, 1991
- [3] : A Hirano & A. Sugiyama : "A Compact Multi-channel Echo Canceller with a Single Adaptive Filter per Channel.". Congrès ISCAS 1992.
- [4] : B. Widrow & S. Stearns : "Adaptive Signal Processing" - Prentice Hall. 1985
- [5] : O. Ait Amrane, E. Moulines, Y. Grenier : "Structure and Convergence Analysis of the Generalized Multi-Delay Adaptive Filter", EUSIPCO'92, Brussels, 1992.