Dix-septième colloque GRETSI, Vannes, 13-17 septembre 1999

Optimisation d'un filtre d'annulation d'écho sous contrôle d'un détecteur de parole locale

Gérard FAUCON, Régine LE BOUQUIN JEANNES

Laboratoire de Traitement du Signal et de l'Image – Université de Rennes 1
Bât. 22, Campus de Beaulieu, 35042 Rennes Cedex, France

Gerard.Faucon@univ-rennes1.fr, Regine.Le-Bouquin-Jeannes@univ-rennes1.fr

Résumé – Ce travail s'inscrit dans le cadre de l'annulation d'écho acoustique pour les télécommunications mains-libres et a pour objectif de prendre en compte le phénomène de double parole pour maintenir les performances de l'annuleur d'écho à un niveau comparable à celui obtenu en présence d'écho seul. Un détecteur basé sur la cohérence partielle permet de décider de la présence du signal désiré et par là-même de bloquer l'adaptation des coefficients du filtre sur ce type de séquence. Une mise en mémoire de ces coefficients permet de prendre en compte le retard à la détection et donne lieu à deux approches voisines. Celles-ci conduisent à un écho résiduel de puissance très inférieure à celle obtenue en entretenant l'adaptation. La perturbation apportée par le signal provenant du locuteur local est annihilée.

Abstract – In the context of hands-free mobile telephony, acoustic echo cancellation (AEC) remains an important problem to deal with. This work addresses the study of an acoustic echo canceller controlled by a near-end speech detector. As a matter of fact, the echo canceller is an adaptive filter whose coefficients are disturbed by the presence of speech coming from the near-end speaker. Consequently, it is useful to stop the adaptation when this signal is detected to get a better estimation of the echo (far-end speaker). In this paper, a near-end speech detector based on the partial coherence - computed from the observations received by two microphones conditioned on the signal emitted by the loudspeaker - is used. Two approaches performing a saving of the AEC coefficients are described. The AEC system controlled by this detector is assessed through the use of the power of the residual echo.

1. Contexte et motivation

La demande sans cesse croissante de systèmes de télécommunications mains-libres stimule les efforts de recherche pour développer des systèmes efficaces, c'est-à-dire transmettant un signal de parole intelligible et de bonne qualité. Dans cette optique, certains problèmes tels que l'annulation d'écho acoustique, la réduction de bruit et la déréverbération doivent être résolus sans altérer la qualité du signal. Le problème traité ici est celui du contrôle d'écho lorsque deux microphones et un haut-parleur sont disponibles. Le signal provenant du locuteur lointain (ou distant) émis par le haut-parleur est reçu, après passage à travers le canal acoustique de la pièce dite locale, sur le microphone et sera donc renvoyé vers ce même locuteur : il est donc nécessaire de procéder à une annulation d'écho. Plus difficile à résoudre en full-duplex en raison de la présence éventuellement simultanée de la parole locale (venant du locuteur proche) et de la parole issue du locuteur lointain (phénomène de double parole), ce problème doit trouver des solutions pertinentes, même si le mode de double parole ne représente que 20% de la longueur d'une communication.

Les annuleurs d'écho sont généralement adaptatifs pour prendre en compte la non-stationnarité du canal acoustique, mais le bruit présent ainsi que la parole locale perturbent l'adaptation. Le bruit est généralement omniprésent et seules des techniques de prétraitement ou d'annulation d'écho robustes au bruit peuvent être envisagées. Le problème de double parole est plus délicat, d'autant que l'énergie de l'écho et de la parole locale sont du même ordre de grandeur.

L'algorithme développé dans ce travail est basé sur la détection de parole locale mise en œuvre à partir de la cohérence partielle et le contrôle de l'adaptation. Deux améliorations sont proposées, prenant en compte le retard à la détection. D'une manière plus générale, le résultat de détection peut être utilisé pour contrôler un système global de réduction de bruit et d'écho [1].

Après avoir donné le principe du détecteur de parole locale et de l'annuleur d'écho étudié, nous montrons comment les performances de la réduction d'écho peuvent être améliorées en optimisant le contrôle de l'AEC par le détecteur et la mise en mémoire des coefficients, performances évaluées en termes de puissance d'écho résiduel.

2. Présentation des modules

2.1 L'annuleur d'écho

Dans cette section, nous présentons l'implémentation du filtre LMS (Least Mean Square) adaptatif conventionnel temporel dans le domaine fréquentiel (Figure 1), qui présente l'avantage de réduire le volume de calculs [2]. Nous résumons ici les étapes importantes de ce traitement. Soit z le signal présent sur le haut-parleur et qui représente l'entrée du filtre LMS adaptatif dont le vecteur de coefficients est noté W .

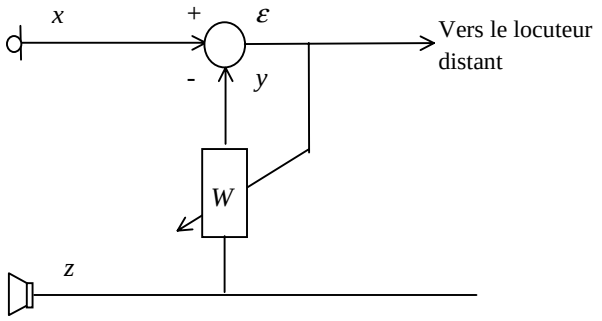

FIG. 1. Filtre LMS adaptatif

La variable x représente l'observation reçue sur le microphone, y est la sortie du filtre et ε la différence entre le signal microphone et le signal estimé. Afin d'implémenter la convolution linéaire des coefficients de pondération du filtre avec son entrée dans le domaine fréquentiel, via la technique "overlap-save", les coefficients sont complétés avec n zéros, et les FFT sont réalisées sur $2n$ points. En pratique, nous notons ω_k le vecteur fréquentiel de longueur $2n$, dont les éléments sont les transformées de Fourier des n coefficients du vecteur temporel sur le k ème bloc, complétés de n zéros,

$$\omega_k^T = \mathrm{FFT}\left[W_k^T\, 0...0\right] \tag{1}$$

De la même manière, Z_k est un vecteur de $2n$ points correspondant à la transformée de Fourier des $(k-1)$ ème et k ème blocs d'entrée :

$$\mathrm{Z}_k^T = \mathrm{FFT}\left[z_{(k-1)n}...z_{kn-1}z_{kn}...z_{kn+n-1}\right]. \tag{2}$$

La sortie du filtre au k ème bloc est obtenue ainsi :

$$\left[y_{kn}...y_{kn+n-1}\right]^T = n \text{ derniers termes de } \mathrm{FFT}^{-1}\left\{\omega_k \otimes \mathrm{Z}_k\right\} \tag{3}$$

où $\otimes$ dénote la multiplication élément par élément de deux vecteurs.

L'équation de mise à jour du vecteur de coefficients dans le domaine fréquentiel s'écrit :

$$\omega_{k+1} = \omega_k + 2\mu\mathrm{FFT}\left[\nabla_k^T 0...0\right]^T, \tag{4}$$

∇_k est le vecteur gradient calculé comme suit :

$$\nabla_k = n \text{ premiers termes de } \mathrm{FFT}^{-1}\left\{E_k \otimes \overline{\mathrm{Z}}_k\right\} \tag{5}$$

(la barre sur Z_k traduit l'opération de conjugaison), avec :

$$E_k = \mathrm{FFT}\left[0...0x_{kn} - y_{kn}...x_{kn+n-1} - y_{kn+n-1}\right]^T. \tag{6}$$

2.2 Le détecteur de parole locale

2.2.1 Présentation

Le détecteur que nous proposons est basé sur la cohérence partielle et suppose que nous ayons à notre disposition deux microphones et un haut-parleur. Les microphones sont suffisamment éloignés pour que les bruits recueillis soient décorrélés. A partir des observations reçues, on calcule la cohérence partielle conditionnellement au signal émis par le haut-parleur. Cette quantité traduit la cohérence entre les deux observations lorsque l'on retranche de celles-ci les grandeurs corrélées au haut-parleur [3]. En conséquence, elle est théoriquement nulle en la seule présence de l'écho et du bruit. Comparée à un seuil au cours du temps, elle permet donc d'attester de la présence éventuelle du signal local.

2.2.2 Implémentation

La cohérence partielle est simple à mettre en œuvre, puisqu'elle s'exprime directement en fonction des cohérences calculées deux à deux à partir des signaux présents sur les microphones et le haut-parleur. En effet, soient x_1, x_2 les observations reçues sur les deux microphones et z le signal présent sur le haut-parleur ; alors, à chaque fréquence f_j la cohérence partielle entre x_1 et x_2 conditionnellement à z s'écrit :

$$\rho_{x_1x_2/z}(f_j) = \frac{\rho_{x_1x_2}(f_j) - \rho_{x_1z}(f_j)\rho_{zx_2}(f_j)}{\sqrt{1-\left|\rho_{x_1z}(f_j)\right|^2}\sqrt{1-\left|\rho_{x_2z}(f_j)\right|^2}} \tag{7}$$

où la cohérence ordinaire calculée entre deux quantités u et v, $\rho_{uv}(f_j)$, se définit comme le rapport de l'interspectre correspondant sur le produit des racines carrées des densités spectrales de puissance (dsp) associées à chaque composante.

En pratique, les dsp sont estimées en utilisant une fenêtre rectangulaire et un recouvrement de 50% :

$$\hat{\gamma}_{uv}(f_j,k) = \frac{1}{K}\sum_{l=k-K+1}^{k} U(f_j,l)\overline{V}(f_j,l) \tag{8}$$

où $U(f_j,l)$ et $V(f_j,l)$ représentent les transformées de Fourier des signaux u et v présents sur un bloc l de N échantillons, k est le bloc courant et K le nombre de blocs sur lequel est faite l'estimation (en l'occurrence $N=256$ et $K=19$ ce qui correspond à 10 blocs adjacents). La cohérence partielle est ainsi calculée sur chaque bloc k. Il a été montré que le biais du module de la cohérence (qu'elle soit ordinaire ou partielle) ne dépend que de ce module [4] et qu'il est d'autant plus important que ce module est faible. Dans [3], nous avons proposé un algorithme qui réduit le biais à chaque fréquence afin d'augmenter la dynamique des valeurs de cohérence. *De facto*, nous obtenons une grandeur corrigée du module de la cohérence partielle à chaque fréquence, notée $\left|\hat{\rho}^c_{x_1x_2/z}(f_j,k)\right|$.

Ensuite, nous calculons la quantité suivante :

$$\overline{\rho}_{x_1x_2/z}(k) = \frac{1}{\frac{N}{2}+1}\sum_{j=1}^{\frac{N}{2}+1}\left|\hat{\rho}^c_{x_1x_2/z}(f_j,k)\right|. \tag{9}$$

Ce moyennage permet une interprétation plus aisée et réduit la variance de l'estimateur. La cohérence partielle corrigée moyennée est comparée à un seuil afin de décider ou non de la présence du signal utile et ce à chaque bloc considéré.

3. Principe et améliorations

Nous donnons ci-après le principe du contrôle ainsi que les améliorations apportées et nous évaluons les performances de l'annuleur d'écho sur signaux simulés. Pour ce faire, on calcule la puissance de l'écho résiduel sur chaque bloc de 128 échantillons. Celui-ci est donné par la différence entre l'écho réel (ici connu) et l'écho estimé. Les différents signaux sont ainsi générés : un bruit blanc gaussien simule le signal haut-parleur z et on en déduit par filtrage les échos reçus par les microphones. De même, on crée deux signaux de parole proche à partir d'un bruit blanc gaussien, indépendant de z. Les bruits reçus par chaque microphone sont des bruits blancs gaussiens et indépendants. Nous construisons la séquence suivante : "bruit, bruit + écho, bruit + écho + signal, bruit + signal". Chaque séquence contient 10 000 échantillons, ce qui correspond pratiquement à 78 blocs de 256 échantillons recouverts à 50%. Les rapports signal à bruit et écho à bruit sur chacun des microphones sont fixés à 6 dB.

3.1 Contrôle de l'AEC par le détecteur de parole locale

Nous savons que la sortie de l'annuleur d'écho commande l'adaptation de ses coefficients. Lorsque le signal proche est présent, celui-ci va donc agir sur l'adaptation et perturber l'identification du canal acoustique. L'idée consiste à bloquer l'adaptation lorsque la parole locale est détectée. Ce blocage sur la durée du signal supposée connue (Figure 2.a) permet de retrouver une puissance d'erreur comparable à celle obtenue après convergence de l'algorithme en absence de signal. Par contre, si l'adaptation est entretenue (Figure 2.b), la puissance de l'erreur est fortement accrue. Après insertion du détecteur pour commander l'adaptation, le résultat obtenu (Figure 2.c) est essentiellement fonction du retard à la détection mais conduit à une puissance d'erreur voisine de celle obtenue en supposant le détecteur parfait. Les puissances sont évidemment nulles en absence d'écho (début et fin de la séquence).

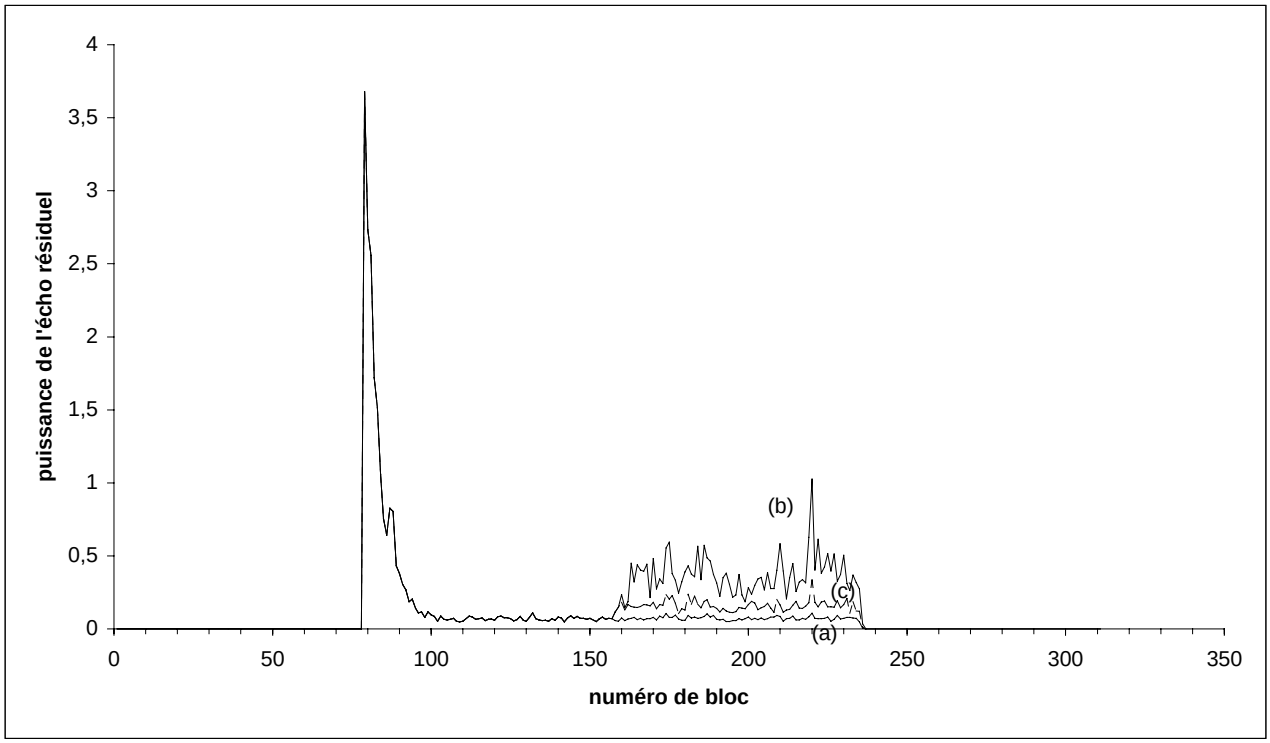

FIG. 2. (a) adaptation bloquée avec un détecteur idéal, (b) adaptation continue, (c) adaptation bloquée avec le détecteur construit

3.2 Mise en mémoire des coefficients de l'AEC

La première amélioration proposée consiste à prendre en compte le retard à la détection ; en effet, suite à un grand nombre de simulations, il apparaît que notre détecteur peut réagir avec un retard de deux blocs au maximum par rapport à une détection idéale. Avec le détecteur réel (Figure 3.a), nous obtenons une puissance d'erreur un peu plus importante qu'avec un détecteur idéal. L'idée que nous exploitons maintenant est, lorsqu'il y a détection, de reprendre les coefficients du filtrage qui ont été appris quelques blocs auparavant lorsque la parole est absente. Il s'agit alors de garder en mémoire les coefficients du filtre au cours du temps et, lorsque le détecteur indique la présence du signal désiré au block k, de réinitialiser les coefficients avec ceux mis en mémoire au bloc $k-2$ obtenus après mise à jour. Ainsi, si la puissance de l'écho résiduel est augmentée sur les blocs k -2 et k -1, elle retrouve, dès le bloc k, une valeur comparable à celle obtenue avant l'arrivée du signal (Figure 3.b). Notons que sur la Figure 3 seule la puissance de l'erreur après le régime transitoire est représentée (en l'occurrence, au-delà du $100^{ème}$ bloc).

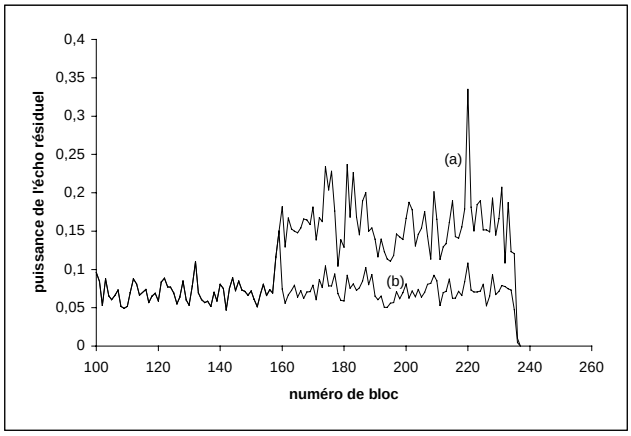

FIG. 3. (a) adaptation bloquée avec le détecteur présenté en 3.1, (b) adaptation bloquée avec mise en mémoire des coefficients du filtre

3.3 Copie de l'annuleur d'écho

L'approche précédente permet de retrouver, dès détection du signal de parole proche, une puissance d'erreur identique à celle obtenue avant l'apparition du signal (Figure 3.b).

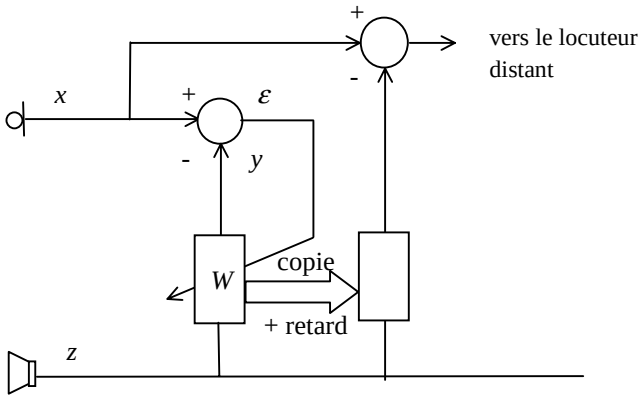

FIG. 4. Filtre avec recopie

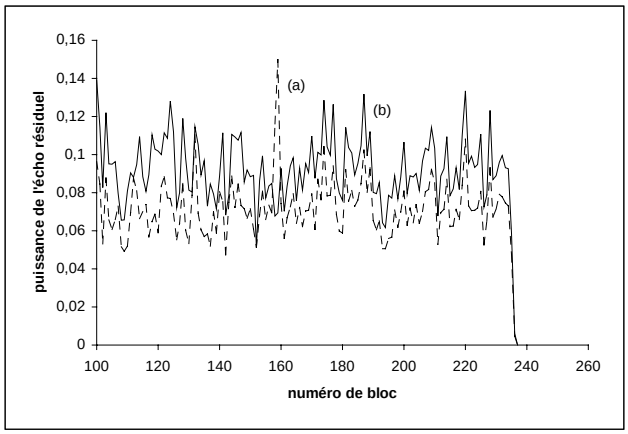

FIG. 5. (a) adaptation bloquée avec le détecteur présenté en 3.2, (b) adaptation bloquée avec recopie du filtre et décalage permanent

Apparaît cependant un pic de puissance sur quelques blocs qui peut être préjudiciable au niveau de l'écoute. En s'inspirant de l'approche et des résultats précédents, une seconde approche est proposée, dans laquelle le filtre adaptatif est recopié à chaque bloc dans un autre filtre avec deux blocs de retard (Figure 4). Ce deuxième filtre construit un écho estimé alors soustrait de la voie microphone. Ainsi, lors des premiers blocs correspondant à l'apparition du signal, ce filtre qui utilise d'anciennes valeurs fournit une estimée correcte de l'écho. Ceci exige simplement que le canal acoustique soit lentement variable. La courbe présentée sur la Figure 5.b montre la puissance de l'écho résiduel obtenue avec cette approche, puissance qui ne présente pas de pic (par opposition à la Figure 5.a) et qui reste très voisine de celle obtenue avec un détecteur idéal.

4. Etude sur signaux réels

Après une campagne de mesure sur signaux simulés, des tests ont été conduits sur signaux réels. Les deux microphones sont espacés de 40 cm pour obtenir des bruits faiblement corrélés. Le bruit ambiant provient d'une voiture roulant à 130 km/h. Les signaux utiles sont enregistrés indépendamment du signal émis sur le haut-parleur, des échos et des bruits, afin de pouvoir faire varier les rapports signal à bruit et écho à bruit. En situation d'adaptation continue, la puissance de l'écho résiduel en sortie du système présente des fluctuations importantes atteignant des valeurs jusqu'à 10 fois celles obtenues lorsque l'adaptation est bloquée en utilisant le détecteur et les systèmes de mise en mémoire.

5. Conclusion

Cette étude montre l'intérêt d'un détecteur de signal associé à un système de mise en mémoire des coefficients pour contrôler l'annuleur d'écho. Celui-ci présente alors en mode de double parole des performances comparables à celles obtenues en absence de signal utile. Notons également que la cohérence ordinaire permet d'obtenir un détecteur d'activité vocale (locuteur proche ou locuteur distant) et que la différence entre cohérence ordinaire et cohérence partielle permet de décider de la présence d'écho. Ainsi, l'exploitation de ces diverses grandeurs permet de savoir quel type de séquence est présent et peut être mise à profit pour d'autres traitements de rehaussement de la parole.

Références

[1] R. Le Bouquin Jeannès, G. Faucon, B. Ayad, "How to improve acoustic echo and noise cancelling using a single talk detector", *Speech Communication*, 20, 3-4, pp. 191-202, 1996.

[2] E.R. Ferrara, "Fast implementation of LMS adaptive filters", *IEEE ASSP*, 28, 4, pp. 474-475, 1980.

[3] G. Faucon, R. Le Bouquin Jeannès, "Double-talk detection with application to acoustic echo cancellation", *ICSPAT*, Boston, pp. 1392-1396, 1998.

[4] G.C. Carter, C.H. Knapp, A.H. Nutall, "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing", *IEEE Trans. on Audio and Electroacoustics*, vol. AU-21, n°4, pp. 337-344, Aug. 1973.