

# ANALYSE DES PERFORMANCES D'UNE NOUVELLE STRUCTURE D'AEC DANS LE DOMAINE TATOUE

I. Marrakchi-Mezghani<sup>(1)</sup>, M. Turki-Hadj Alouane<sup>(1)</sup>, S. Djaziri-Larbi<sup>(1)</sup>, M. Jaïdane-Saïdane<sup>(1)</sup>, G. Mahé<sup>(2)</sup>

<sup>(1)</sup> Unité Signaux et Systèmes, Ecole Nationale d'Ingénieurs de Tunis, Tunisia  
email: mezghani\_imen@yahoo.fr, {m.turki, sonia.larbi, meriem.jaidane}@enit.rnu.tn

<sup>(2)</sup> CRIP5, Université René Descartes, Paris V, France  
email: gael.mahé@math-info.univ-paris5.fr

## ABSTRACT

Une nouvelle approche d'annulation adaptative de l'écho acoustique a été présentée dans [1]. Le Watermarked AEC (WAEC) proposé, exploite les bonnes propriétés statistiques (bancheur et stationnarité) du bruit blanc inséré par tatouage audio dans le signal de parole source de l'écho acoustique. Dans ce papier, nous présentons une analyse approfondie des bonnes propriétés du WAEC en terme de vitesse et de finesse de convergence. La nette supériorité du WAEC par rapport à un AEC classique est due au fait que les signaux impliqués dans le fonctionnement du WAEC sont très peu corrélés et beaucoup plus stationnaires.

## 1. INTRODUCTION

La tendance actuelle, en tatouage audio-numérique, est l'insertion imperceptible dans le signal audio d'une information numérique utile. Jusqu'à présent, les recherches dans ce domaine se sont principalement focalisées sur des objectifs de maximisation du débit du flux binaire inséré et d'optimisation du procédé d'extraction. Toutefois dans certains travaux [2, 3], le tatouage est utilisé pour améliorer les caractéristiques des signaux traités dans différentes applications audio dont l'annulation d'écho acoustique (AEC). Le Watermarked AEC (WAEC) est basé sur le concept d'insertion, dans le signal de parole, d'un bruit blanc non audible qui va s'imprégner de l'information relative au chemin acoustique à identifier. Ainsi, dans le WAEC, le filtre adaptatif estimant le chemin acoustique traite des signaux beaucoup plus stationnaires et décorrés que ceux traités par le filtre adaptatif d'un AEC classique. Affranchi de la forte corrélation et de la non stationnarité de la parole, le WAEC présente des performances nettement meilleures que celles d'un AEC classique. Dans [1], on s'est focalisé sur la présentation du concept du système WAEC. Dans le cadre de ce papier, nous nous intéressons à l'analyse approfondie des performances du WAEC proposé.

L'étude proposée repose sur une analyse des propriétés statistiques des signaux impliqués dans le fonctionnement du filtre adaptatif estimant l'écho à travers la caractérisation de l'évolution temporelle de la déviation mesurant la différence entre le chemin d'écho estimé et celui réel. En particulier, nous nous sommes focalisés sur l'évaluation des degrés de corrélation et de stationnarité de ces signaux. La mesure de ces derniers est faite par les indices de stationnarités issus d'une analyse temps/fréquence [4, 5] [4, 5]. En fait, la forte non stationnarité des signaux affecte les performances de l'algorithme NLMS utilisé pour adapter le filtre estimant l'écho en régime permanent. Il a été alors montré

que dans les signaux impliqués dans l'adaptation du filtre NLMS du WAEC, sont beaucoup plus stationnaires que ceux utilisés par le filtre du AEC classique. Cet avantage a constitué un facteur important dans la supériorité en régime permanent du WAEC sur le AEC classique. Ceci a été mis en évidence par l'étude des variations temporelles de la déviation quadratique moyenne.

L'étude du comportement moyen de la déviation, a montré que la faible corrélation des signaux impliqués dans le fonctionnement WAEC a permis une accélération remarquable de la vitesse de convergence du filtre adaptatif du WAEC comparé à celle du filtre adaptatif de l'AEC classique.

Le papier est structuré comme suit : dans la section 2 nous présentons la structure du WAEC proposé. Une mise en évidence du contexte stationnaire du WAEC est présentée dans la section 3. Les sections 4 et 5 confirment l'amélioration des comportements moyen et quadratique moyen du WAEC.

## 2. PRÉSENTATION DU WAEC

L'idée de départ dans la conception du WAEC était de piloter l'AEC par un signal stationnaire dans le but d'en améliorer les performances transitoires et permanentes comparées à celles obtenues dans le cas d'un AEC classique avec entrée tatouée [3].

La structure de l'AEC classique est modifiée de manière à ce que l'écho soit estimé en utilisant le signal de tatouage à son entrée, ce dernier étant nettement plus stationnaire et plus blanc que le signal original. Ce signal est donc considéré porteur de la réponse impulsionnelle acoustique.

Nous utilisons pour le masquage psychoacoustique du tatouage un filtre de mise en forme perceptif dont la réponse fréquentielle est obtenue à partir de l'enveloppe spectrale du signal de parole  $x_n$ . L'introduction de ce filtre prédictif de la parole permet le masquage perceptif de la séquence  $w_n$  ainsi que la décorrélation du signal d'entrée tatoué du premier filtre adaptatif (de type NLMS), dont elle accélère la convergence. La nouvelle structure de l'AEC proposé (WAEC) est donnée par la Fig. 1.

Le signal de parole pouvant être considéré comme un processus autorégressif à l'ordre  $M$ , on a :

$$x_n = (P_n)^T X_{n-1} + e_n^x,$$

où  $e_n^x$  est l'erreur de prédiction du signal  $x_n$  de variance  $\sigma_{e_n^x}^2(n)$  et  $X_{n-1} = (x_{n-1}, \dots, x_{n-M})^T$  est le vecteur observation.  $P_n = (p_n^1, \dots, p_n^M)^T$  représente le vecteur des coefficients du filtre prédictif du signal d'entrée  $x_n$ , dont l'adaptation est

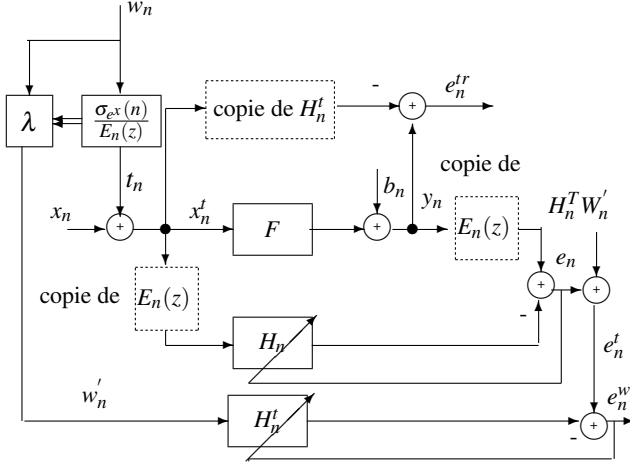


FIG. 1 – Schéma de l'AEC proposé (WAEC)

faite par l'algorithme NLMS. La fonction de transfert du filtre blanchisseur est donc  $E_n(z) = 1 - P_n(z)$ .

Ainsi, le filtre de mise en forme perceptif utilisé ici est adaptatif et de fonction de transfert :  $S_n(z) = \frac{\sigma_{e^x}(n)}{E_n(z)}$ .

Des tests subjectifs d'évaluation de la qualité audio après tatouage ont révélé qu'une atténuation supplémentaire de  $\lambda = 13$  dB de la réponse fréquentielle  $S_n(f)$  est nécessaire pour assurer l'inaudibilité du signal de tatouage  $t_n$ .

On désigne par la suite  $e_n^x$ ,  $e_n^{x'}$  et  $e_n^b$  les sorties du filtre  $E_n(z)$  pour les entrées  $x_n$ ,  $x_n^t$  et  $b_n$  respectives (où  $b_n$  est un bruit blanc, centré et indépendant des signaux  $x_n^t$  et  $t_n$ ).

Le signal d'écho de référence du premier filtre  $H_n$  est donné par :

$e_n^y = F^T E_n^x + F^T W_n^t + e_n^b$ , où  $E_n^x = (e_n^x, \dots, e_{n-p+1}^x)^T$ ,  $F$  est le chemin d'écho à identifier tronqué à l'ordre  $p$  et :

$$W_n^t = \lambda \sigma_{e^x}(n) W_n.$$

L'erreur  $e_n$  qui contrôle l'adaptation du filtre  $H_n$  est  $e_n = e_n^y - H_n^t (E_n^x + W_n^t)$  et l'écho de référence pour le second filtre adaptatif  $H_n^t$  devient :

$$e_n^t = F^T W_n^t + \underbrace{e_n^b + V_n^T E_n^x}_{e_n^b}, \quad (1)$$

où  $V_n = F - H_n$  et  $e_n^b$  est un bruit non stationnaire, puisqu'il est lié au signal non stationnaire  $e_n^x$ . La variance de ce bruit est  $\sigma_{e_n^b}^2(n) = E[(e_n^b)^2]$ . L'expression de l'erreur  $e_n^t$  donnée par la relation (1) montre que l'information utile pour le filtre  $H_n^t$  est le terme  $F^T W_n^t$  qui représente la convolution entre le chemin d'écho à identifier et une séquence nettement plus blanche et plus stationnaire que le signal de parole. C'est ainsi que le signal  $w_n^t$  sera utilisé comme entrée du filtre  $H_n^t$  évoluant comme suit :

$$H_{n+1}^t = H_n^t + \mu_n^t e_n^w W_n^t, \quad (2)$$

où l'erreur  $e_n^w$  est donné par :

$$e_n^w = (V_n^t)^T W_n^t + e_n^b, \quad (3)$$

avec  $V_n^t = F - H_n^t$ . Le signal d'écho résiduel qui sera réellement transmis est :

$$e_n^{tr} = y_n - (H_n^t)^T X_n^t.$$

Ainsi, les performances de  $H_n^t$  seront comparées dans la suite à celles du filtre adaptatif,  $H_n$ , piloté par le NLMS, dont l'entrée est le signal  $x_n^t$  et le signal de référence est  $y_n$ . Ainsi, l'écho résiduel est donné par :

$$e_n = y_n - H_n^T X_n^t,$$

qui est à comparer à l'écho résiduel transmis  $e_n^{tr}$ .

Dans cette section on a présenté la structure du WAEC. Dans ce qui suit, on s'intéresse à l'évaluation des performances du WAEC en régimes transitoire et permanent. Sachant que l'étude théorique du comportement quadratique moyen est difficile à mener dans le cas du NLMS à cause de la corrélation et de la non stationnarité des signaux traités, l'analyse suivante sera basée essentiellement sur des résultats de simulation obtenus sous les conditions suivantes : les signaux de parole sont échantillonnés à 8 kHz et le chemin d'écho à identifier correspond à la réponse impulsionnelle d'un habitacle de voiture tronquée à 200 coefficients. Le rapport signal à bruit, ( $RSB = 10 \log_{10}(\frac{E(x_n^2)}{E(b_n^2)})$ ), est de 40 dB. Les ordres des filtres  $P$  et  $F$  ont été fixés respectivement à  $M = 4$  et  $p = 200$ . Pour les deux filtres couplés du WAEC, nous prenons le même pas d'adaptation  $\mu^t$  et le même ordre  $p$ .

### 3. MISE EN ÉVIDENCE DU CONTEXTE PLUS STATIONNAIRE DU WAEC

Les comportements en régimes transitoire et permanent du WAEC et de l'AEC classique sont étudiés respectivement à travers les variations temporelles des vecteurs déviations  $V_n^t$  et  $V_n = F - H_n$ .

Dans le cas de l'AEC classique, il est aisé de montrer que :

$$V_{n+1} = (I - \mu \frac{X_n^t (X_n^t)^T}{\|X_n^t\|^2}) V_n + \mu b_n \frac{X_n^t}{\|X_n^t\|^2}. \quad (4)$$

Dans le cas du WAEC et en combinant (1), (2) et (3), on obtient la relation :

$$V_{n+1}^t = (I - \mu^t \frac{W_n^t (W_n^t)^T}{\|W_n^t\|^2}) V_n^t + \mu^t e_n^b \frac{W_n^t}{\|W_n^t\|^2}. \quad (5)$$

D'après (4) et (5), quatre signaux sont impliqués dans l'évolution de  $V_n$  et de  $V_n^t$ , à savoir  $x_n^t$  et  $w_n^t$  d'une part, et  $S1_n$  et  $S2_n$  d'autre part, avec :

$$S1_n = \frac{b_n \times x_n^t}{\|x_n^t\|^2} \text{ et } S2_n = \frac{e_n^b \times w_n^t}{\|w_n^t\|^2}.$$

Vu que l'algorithme NLMS est conçu pour bien fonctionner dans un contexte stationnaire, les performances des filtres  $H_n$  et  $H_n^t$  seront étroitement liées à la stationnarité de ces quatre signaux.

L'évaluation du degré de stationnarité d'un signal peut être faite par l'outil indice de stationnarité [4, 5]. Fondé sur une analyse temps/fréquence, cet indice détecte les transitoires rapides des signaux. Parmi les trois indices présentés dans [4, 5], il a été mis en évidence que les indices de Küllback

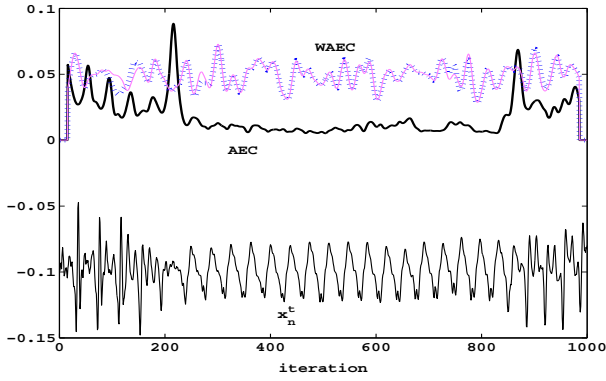


FIG. 2 – Indices de Bhattacharyya pour un extrait de  $x_n^t$  (en noir), pour le signal  $w_n'$  correspondant (en bleu) et pour  $w_n$  (en violet).

et de Bhattacharyya sont les plus adaptés aux mesures des variations rapides propres aux signaux audio. Nous avons donc choisi d'utiliser l'indice de Bhattacharyya comme indice de stationnarité. L'indice de Bhattacharyya à un instant donné  $n$  exprime la distance entre deux imagettes  $I_1(n; \tau, f)$  et  $I_2(n; \tau, f)$  de la représentation temps/fréquence (RTF) de part et d'autre de cet instant  $n$ , avec :

$$\begin{aligned} I_1(n; \tau, f) &= RTF(n-L+\tau, f) \\ I_2(n; \tau, f) &= RTF(n+\tau, f) \end{aligned} \quad (6)$$

où  $L$  est la largeur des deux imagettes (en temps),  $\tau$  est dans  $[0, L]$  et  $f$  est la fréquence.

La valeur de cet indice est donc donnée par :

$$SI(n) = -\log \left( \int_{\tau=0}^L \int_{-\infty}^{+\infty} \sqrt{NI_1(n; \tau, f) NI_2(n; \tau, f)} df d\tau \right) \quad (7)$$

où  $NI_{k(k=1,2)}$  est l'imagette  $I_k$  normalisée.

L'apparition d'un pic dans l'évolution de  $SI(n)$  indique une variation rapide du spectre du signal, témoignant ainsi de la présence d'une zone de transitoire.

La Fig. 2 présente l'évolution dans le temps des indices  $SI_{x^t}$ ,  $SI_{w'}$  et  $SI_w$  correspondant respectivement aux signaux  $x_n^t$ ,  $w_n'$  et  $w_n$ . Ces indices sont calculés sur des trames successives de 128 échantillons. Cette figure montre que  $SI_{x^t}$  présente plus de pics que  $SI_{w'}$  et  $SI_w$ . On remarque aussi que d'une part, les indices  $SI_{w'}$  et  $SI_w$  sont quasi superposés, et que d'autre part, l'indice de stationnarité d'un bruit blanc présente des variations faibles mais non nulles. Ainsi, on déduit que le signal  $w_n'$  est nettement plus stationnaire que  $x_n^t$ .

La Fig. 3 illustre l'évolution des indices  $SI_{S1}$  et  $SI_{S2}$  relatifs aux signaux  $S1_n$  et  $S2_n$ . Cette figure montre que le nombre de pics caractérisant l'évolution de  $SI_{S1}$  est nettement plus important que celui mis en évidence par l'évolution de  $SI_{S2}$ . Ainsi, selon cette simulation, le signal  $S2_n$  est nettement plus stationnaire que  $S1_n$ . Pour cette simulation les pas d'adaptation  $\mu$  et  $\mu'$  sont fixés à 0.02. Ces mesures comparatives des indices de stationnarité des quatre signaux, impliqués dans l'évolution des filtres adaptatifs, permettent de conclure que le filtre adaptatif  $H_n^t$  fonctionne dans un environnement plus stationnaire que celui de  $H_n$ . Ainsi, on s'attend à ce que le filtre adaptatif du WAEC proposé présente en particulier en régime permanent de meilleures performances que le filtre

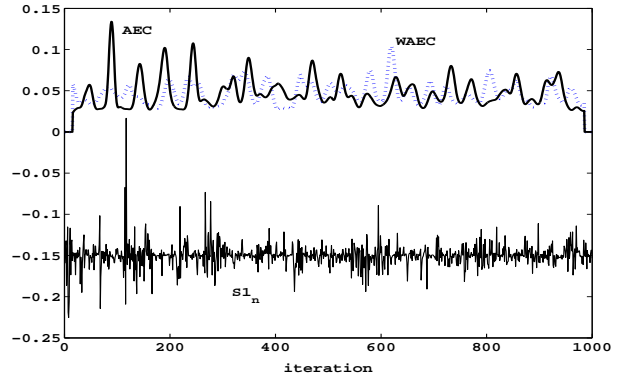


FIG. 3 – Indices de Bhattacharyya pour le signal  $S1_n$  (en noir) et pour le signal  $S2_n$  (en bleu).

adaptatif du AEC classique. La stationnarisation du contexte, nous permettra de s'affranchir de la mauvaise influence de la non stationnarité du signal de parole sur les performances du filtre adaptatif estimant l'écho.

#### 4. PERFORMANCES EN RÉGIME TRANSITOIRE DU WAEC

Le comportement en régime transitoire est classiquement étudié à travers la déviation moyenne instantanée. En particulier, nous nous intéressons à la comparaison des vitesses de convergence du WAEC et du AEC classique. Sachant que  $b_n$  est centré et sous l'hypothèse classique d'indépendance entre  $V_n$  et  $X_n^t$ , il est aisé de déduire respectivement à partir des relations (4) et (5) que :

$$E[V_{n+1}] = (I - \mu R_{x^t}(n))E[V_n], \quad (8)$$

et

$$E[V_{n+1}^t] = (I - \mu' R_{w'}(n))E[V_n^t], \quad (9)$$

$$\text{où } R_{x^t}(n) = E \left[ \frac{X_n^t (X_n^t)^T}{\|X_n^t\|^2} \right] \text{ et } R_{w'}(n) = E \left[ \frac{W_n' (W_n')^T}{\|W_n'\|^2} \right].$$

A partir des équations (8) et (9), on remarque que la vitesse de convergence des filtres  $H_n$  et  $H_n^t$  est intimement liée au conditionnement des matrices d'autocorrélation respectives  $R_{x^t}(n)$  et  $R_{w'}(n)$ .

Il est à noter que le conditionnement est d'autant plus élevé que le signal est corrélé. Comme le montre la Fig. 4, la matrice  $R_{w'}(n)$  est nettement mieux conditionnée que  $R_{x^t}(n)$ , ce qui s'explique par la corrélation moindre de  $w_n'$ . Ce résultat permet de prévoir une convergence plus rapide du filtre  $H_n^t$ .

#### 5. PERFORMANCES EN RÉGIME PERMANENT DU WAEC

L'analyse des performances en régime permanent est faite à travers la déviation quadratique moyenne instantanée définie par  $DQM(n) = E[V_n^T V_n]$ .

A partir des relations(4) et (5), il est aisé de montrer les deux

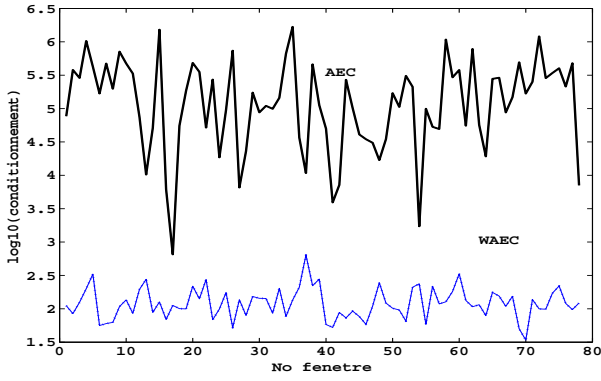


FIG. 4 – Conditionnement de  $R_w'(n)$  (en bleu) et de  $R_x'(n)$  (en noir).

relations de récurrence suivantes :

$$E[\|V_{n+1}\|^2] = \underbrace{\mu^2 \sigma_b^2 E\left[\frac{1}{\|X_n^t\|^2}\right]}_{P_n^K} + E[V_n^T (I - \mu(2 - \mu) \frac{X_n^t (X_n^t)^T}{\|X_n^t\|^2}) V_n]; \quad (10)$$

$$E[\|V_{n+1}^t\|^2] = \underbrace{(\mu^t)^2 \sigma_{e^{b'}}^2(n) E\left[\frac{1}{\|W_n^t\|^2}\right]}_{P_n^D} + E[(V_n^t)^T (I - \mu^t(2 - \mu^t) \frac{W_n^t (W_n^t)^T}{\|W_n^t\|^2}) V_n^t], \quad (11)$$

où  $\sigma_b^2$  et  $\sigma_{e^{b'}}^2(n)$  représentent les variances respectives de  $b_n$  et de  $e_n^{b'}$ .

D'après les relations (10) et (11), l'évolution temporelle des termes  $P_n^K$  et  $P_n^D$  influe sur la nature de l'évolution et de la valeur moyenne de la déviation quadratique ( $DQM$ ) respectives des filtres  $H_n$  et  $H_n^t$ , une fois le régime permanent est établi. Il est à noter que les termes  $P_n^K$  et  $P_n^D$  sont étroitement liés aux signaux  $S1_n$  et  $S2_n$  étudiés dans la section 3. Ainsi, les variations temporelles de  $P_n^K$  et  $P_n^D$  sont conditionnées respectivement par la stationnarité des signaux  $S1_n$  et  $S2_n$  qui eux même sont liés à la stationnarité des signaux  $x_n^t$  et  $w_n^t$ . En effet, plus ces quatre signaux sont stationnaires, plus les variations de  $P_n^K$  et  $P_n^D$  sont douces.

Sur la Fig. 5, illustrant les variations temporelles de ces termes pour  $\mu = \mu^t = 0.02$ , on remarque que :

- l'évolution de  $P_n^K$  est caractérisée par des pics de fortes amplitudes. On peut donc s'attendre à des variations brutales au niveau de l'évolution de la  $DQM(n)$  en régime permanent
- la valeur moyenne du terme  $P_n^K$  est supérieure à celle de  $P_n^D$ . Ceci permet de prévoir des variations plus faibles de la  $DQM(n)$  relative au WAEC.

La Fig. 6 montre les évolutions temporelles des  $DQM(n)$  du WAEC et de l'AEC classique obtenues respectivement pour  $\mu^t = 0.02$  et  $\mu = 0.04$ . On remarque que :

- La vitesse de convergence de  $H_n^t$  est plus élevée que celle du filtre  $H_n$ , bien que  $\mu^t$  soit inférieur à  $\mu$ . Ce résultat confirme les résultats de l'analyse élaborée dans la Section 4.
- une fois le régime permanent établi, les variations de la  $DQM$  du filtre  $H_n^t$  sont nettement plus lentes et plus faibles que celles de  $H_n$ .

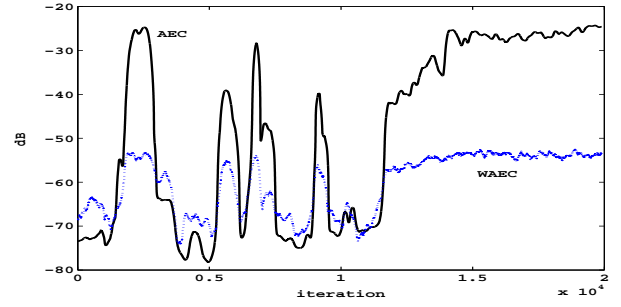


FIG. 5 – Variations temporelles de  $P_n^D$  (en bleu) et de  $P_n^K$  (en noir). Pas d'adaptation :  $\mu = \mu^t = 0.02$

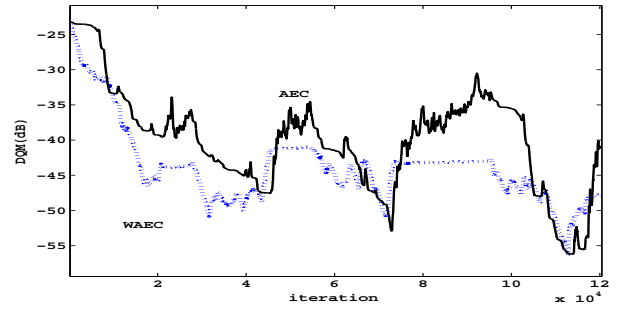


FIG. 6 – Variations temporelles de la  $DQM$  pour le WAEC (en bleu) et pour l'AEC classique (en noir).

Pas d'adaptation :  $\mu = 0.04$  et  $\mu^t = 0.02$

## 6. CONCLUSION

Nous avons présenté dans ce papier une analyse comparative des performances du WAEC et du AEC classique.

Cette étude, basée essentiellement sur des résultats de simulation, nous a permis de mettre en évidence que le WAEC fonctionne dans un contexte nettement plus stationnaire et moins corrélé que celui dans lequel fonctionne le AEC classique. Ainsi, des résultats convaincants en terme de vitesse de convergence et de finesse de convergence ont été obtenus par le WAEC proposé (un gain de 10 dB et plus au niveau de l'ERLE<sup>1</sup> est atteint).

## REFERENCES

- [1] I. Marrakchi, M. Turki, S. Larbi, M. Jaïdane et G. Mahé, "Speech processing in the watermarked domain : application in adaptive acoustic echo cancellation", *soumis à EUSIPCO*, 2006.
- [2] A. Gilloire, V. Turbin, "Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers", in *ICASSP*, Seattle, USA, 1998.
- [3] S. Larbi, M. Jaïdane, M. Turki et M. Bonnet, "Réflexion sur un annuleur d'écho robuste aux non stationarités de la parole", in *CORESA*, Compression et représentation des signaux audio visuels, Dijon, France, 2001.
- [4] H. Laurent and C. Doncarli, "Stationarity index for abrupt changes detection in the time frequency plane", *IEEE Signal Processing Letters*, vol. 5, no. 2, pp.43 -45, 1998.
- [5] S. Larbi and M. Jaïdane, "Audio watermarking : a way to modify audio statistics", *IEEE Trans. on Signal Processing*, vol. 53(2), 2005.

<sup>1</sup>Error Return Loss Enhancement