16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

A TRADE-OFF BETWEEN CONVERGENCE SPEED AND MISADJUSTMENT FOR FILTERING DISCONTINUOUS SPEECH SIGNALS

J.M. Górriz, J. Ramírez, C.G. Puntonet

Dpt. Signal Theory and Communications, ETSIIT, University of Granada, Spain
email: gorriz@ugr.es

S. Cruces-Álvarez,

Dpt. Signal Theory and Communications, University of Seville, Spain
email: sergio@us.es

D. Erdogmus

Dpts. of CSEE and BME Oregon Health and Science University, USA.
email: derdogmus@ieee.org

and E.W. Lang

Ins. Biophysik und physikalische Biochemie, University of Regensburg, Germany.
email: elmar.lang@biologie.uni-regensburg.de.

ABSTRACT

In this paper we propose a novel LMS algorithm in combination with a voice activity detector (VAD) for filtering speech sounds in the Adaptive Noise Cancelation (ANC) problem. The filtering stage is based on the minimization of the squared Euclidean norm of the difference weight vector under a stability constraint over the *a posteriori* estimation error. To this purpose, the Lagrangian methodology has been used in order to propose a non-linear adaptation defined in terms of the product of differential inputs and errors. This approach yields better tracking ability under conditions held in Discontinuous Transmission (DTX) systems than previous approaches. In addition the use of a precise VAD provides two operation modes in order to obtain the best trade-off between misadjustment and convergence speed in speech/non-speech frames. The experimental analysis carried out on the AURORA 3 speech databases provides an extensive performance evaluation together with an exhaustive comparison to standard LMS algorithms including the normalized (N)-LMS, and other recently reported LMS algorithms such as the Modified (M)-NLMS or the Normalized Data Nonlinearity (NDN)-LMS Adaptation.

1. INTRODUCTION

The widely used least-mean-square (LMS) algorithm has been successfully applied to many filtering applications including modeling, equalization, control, echo cancelation biomedicine or beamforming [1]. The typical noise cancelation scheme is shown in Figure 1. Two distant microphones

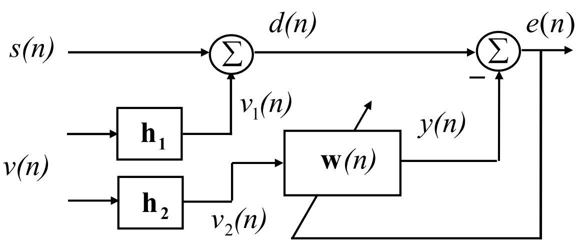

Fig. 1. Adaptive Noise Canceler.

are needed for such application to capture the nature of the noise and the speech sound simultaneously. The correlation between the additive noise that corrupts the clean speech (primary signal) and the random noise in the reference input (adaptive filter input) is necessary to adaptively *cancel* the noise of the primary signal. The adjustable weights are typically determined by the least mean squares (LMS) algorithm [1] because of its simplicity, ease of implementation and low computational complexity. The weight update equation for the adaptive noise canceler (ANC) is:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e^*(n)\mathbf{x}(n) \quad (1)$$

where μ is a step-size parameter, $e^*(n)$ denotes the complex conjugate of output signal $e(n)$, and $\mathbf{x}(n) = (v_2(n), \ldots, v_2(n-L-1))^T$ is the data vector containing L samples of the reference signal $v_2(n)$.

Many ANCs [1, 2, 3] have been proposed in the past years using modified least mean-squares (LMS) algorithms in order to simultaneously improve the tracking ability and speed of convergence. Bershad studied the performance of the Normalized LMS (NLMS) with an adaptive step size in [5] showing advantages in convergence time and steady state. Later, Douglas and Meng [3] proposed the optimum nonlinearity for any input probability density of the independent input data samples, obtaining the Normalized Data Nonlinearity adaptation (NDN-LMS). Although the latter algorithm is designed to improve steady-state performance, its derivation did not consider the ANC in case of a strong target signal in the primary input [2]. Greenberg's Modified-LMS (M-LMS) extended the latter approach to the case of the ANC with the nonlinearity applied to *the data vector and the target signal itself*, obtaining substantial improvements in the performance of the canceler. The disadvantage of this method is that it requires *a priori* information about the processes which is generally unknown. This paper shows a novel adaptation for filtering speech signals in discontinuous speech transmission (DTX) systems, which are characterized by sudden changes of the signal statistics. The method is derived assuming stability in the sequence of *a posteriori* errors instead of the more restrictive hypothesis used in previous approaches [6], that is, enforcing it to vanish. The result of the Lagrange minimization is *the application of the NLMS algorithm to a new set of difference signals that is more suitable for ANC of speech signals in DTX systems*. The combination of the proposed method with an effective VAD [9]-[15] allows to change the operation of the algorithm over speech/non-speech segments which provides a better tracking ability to the adverse non-stationary environment.

2. CONSTRAINED STABILITY (CS)-LMS ALGORITHM

Our approach belongs to the class of LMS algorithms that try to find the optimal weight adaptation scheme [5] rather than optimizing the step size function [2]. The derivation of the nonlinearity applied to the *target signal and the data vector* is achieved by using the Lagrangian formulation relaxing the conditions imposed on the solution unlike the LMS algorithm which enforces the error sequence to vanish [7]. Consider the constrained optimization problem that provides the following cost function:

$$\begin{aligned}&\mathscr{L}(\mathbf{w}(n+1)) = ||\delta\mathbf{w}(n+1)||^2\\ &+Re\left[\lambda^*(e^{[n+1]}(n) - e^{[n+1]}(n-1))\right]\end{aligned} \tag{2}$$

where $\delta\mathbf{w}(n+1) = \mathbf{w}(n+1) - \mathbf{w}(n)$, $e^{[k]}(n) = d(n) - \mathbf{w}^H(k)\mathbf{x}(n)$ is the error at time n obtained from the weight vector at time k, ^{H} denotes the Hermitian and λ is the Lagrange multiplier. The solution to this optimization problem $\mathbf{w}^{opt}(n+1)$ minimizes the norm of the difference between two consecutive weight vectors and satisfies the equilibrium constraint $e^{[n+1]}(n) = e^{[n+1]}(n-1)$ if

$$\lambda = \frac{2\delta e^{[n]}(n)}{||\delta\mathbf{x}(n)||^2} \tag{3}$$

where $\delta e^{[n]}(n) = e^{[n]}(n) - e^{[n]}(n-1)$ is the difference between the *a priori* error sequence and $||\delta\mathbf{x}(n)||^2 = ||\mathbf{x}(n) - \mathbf{x}(n-1)||^2$ is the squared norm of the difference between two consecutive input vectors, subsequently called *difference input vector* for short. This equilibrium constraint imposes *stability* onto the sequence of *a posteriori errors*, i.e. the optimal solution $\mathbf{w}^{opt}(n+1)$ is the one that renders the sequence of errors as smooth as possible. Thus, the minimum of the Lagrangian function satisfies the following update condition:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\delta\mathbf{x}(n)(\delta e^{[n]}(n))^*}{||\delta\mathbf{x}(n)||^2} \tag{4}$$

Thus, the novel condition on the sequence of errors updates the weight vector using the concurrent change of the processing variables which variance is lower if they are strongly correlated, i.e. this operation reduces the variation on the statistics in the channel. The weight adaptation rule can be made more robust by introducing a small number ε into the denominator to prevent numerical instabilities in case of a vanishingly small squared norm $||\delta\mathbf{x}(n)||^2$ and by multiplying the weight increment by a constant step size μ to control the speed of the adaptation. Note that the equilibrium condition enforces the convergence of the algorithm if $||\delta\mathbf{x}(n)||^2 \neq 0$. The convergence of the algorithm, the Excess Minimum Squared Error (EMSE) and the Misadjustment (M) could be readily analyzed in a fashion similar to the NLMS algorithm for the proposed nonlinearities [7] since *the adaptation rule is equivalent except for the difference data*. In addition, under certain conditions imposed onto the difference data and the step size μ of the adaptation rule, the deterministic CS-LMS algorithm converges to the Wiener solution $\mathbf{w}_o$.

3. RELATIONSHIP BETWEEN CS-LMS AND NLMS.

As shown in the previous section, the CS-LMS adaptation rule is equivalent to the NLMS algorithm but over a different set of data. It is interesting to consider the conditions needed for both algorithms to have the same optimal Wiener solution. Let $\mathbf{x}_1(n) = \delta\mathbf{x}(n)$ denote the difference incoming signal, $d_1(n) = \delta d(n) = d(n) - d(n-1)$ the difference desired signal and $e_1(n) = \delta e(n) = d_1(n) - \mathbf{w}^H\mathbf{x}_1$ the error signal. Using this new set of transformed data and applying the Wiener-Hopf methodology, the classical filtering problem is to minimize the cost function:

$$\mathbf{w}_{o,1} = \arg\min_{\mathbf{w}} E[|e_1(n)|^2] \tag{5}$$

Note that the gradient of the latter cost function defines the deterministic CS-LMS algorithm, that is, in terms of expected values. The optimal Wiener solution for the novel data set is defined to be:

$$\mathbf{w}_{o,1} = \mathbf{R}_{\mathbf{x_1}}^{-1}\mathbf{r}_{\mathbf{d_1x_1}} \tag{6}$$

where the autocorrelation matrix $\mathbf{R}_{\mathbf{x_1}}$ is assumed to be full rank and $\mathbf{r}_{\mathbf{d_1x_1}}$ is the cross correlation vector. Generally, $\mathbf{w}_o \neq \mathbf{w}_{o,1}$ but if the desired signal is assumed to be generated by the multiple linear regression model: $d(n) = \mathbf{w}_o^H\mathbf{x}(n) + e_o(n)$, where $e_o(n)$ is an uncorrelated white-noise process that is statistically independent of the input vector $\mathbf{x}(n)$ then:

$$\mathbf{r}_{\mathbf{d_1x_1}} = \mathbf{R}_{\mathbf{x_1}}\mathbf{w}_o \Rightarrow \mathbf{w}_o = \mathbf{w}_{o,1} \tag{7}$$

4. CONVERGENCE ANALYSIS OF THE CS-LMS ALGORITHM

Firstly, lets consider the convergence of the deterministic CS-LMS algorithm which is described as the Steepest Descent algorithm on the novel data set. From the algorithm adaptation we can formulate the following deterministic adaptation for the CS-LMS algorithm:

$$\begin{aligned}&\mathbf{w}(n+1) = \mathbf{w}(n)+\\ &\hat{\mu}\left[-\nabla\{\,J^{[n]}(n) + J^{[n]}(n-1)\} + 2Re\{\nabla \hat{J}^{[n]}(n)\}\right]\end{aligned} \tag{8}$$

where $J(n) = E[|e(n)^2|]$ and $\hat{J}^{[n]}(n) = E[e^{[n]}(n)(e^{[n]}(n-1))^*]$. In terms of the novel data model $\{\mathbf{x}_1(n), d_1(n), \delta e(n)\}$ introduced in section 3 the adaptation could be expressed as:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \hat{\mu}\left[-\nabla\{\,J_1^{[n]}(n)\}\right] \tag{9}$$

where $J_1^{[n]}(n) = E[|\delta e(n)|^2]$. Since $\delta e(n) = d_1(n) - \mathbf{w}^H\mathbf{x}_1(n)$ then:

$$J_1^{[n]}(n) = \sigma_{d_1}^2 - \mathbf{w}^H\mathbf{r}_{\mathbf{d_1x_1}} - \mathbf{r}_{\mathbf{d_1x_1}}{}^H\mathbf{w} - \mathbf{w}^H\mathbf{R}_{\mathbf{x_1}}\mathbf{w} \tag{10}$$

It is readily seen that necessary and sufficient condition for convergence of this modified algorithm can be expressed, in terms of the step-size parameter $\hat{\mu}$ and the largest eigenvalue λ_{max} of the difference matrix $\Delta\mathbf{R}(1) = \mathbf{R}(0) - \mathbf{R}(1)$ where $\mathbf{R}(k) = E[\mathbf{x}(n+k)\mathbf{x}^H(n)]$, as:

$$0 < \hat{\mu} < \frac{1}{\lambda_{max}} \tag{11}$$

4.1. Convergence Analysis of the stochastic CS-LMS algorithm

From (4) it can be readily seen that the stochastic CS-LMS algorithm is equivalent to the NLMS adaptation over the novel difference data set which was defined in the previous section. Thus, once we have proven that they have the same Wiener solution and the stochastic CS-LMS is equivalent to the NLMS on the novel data set, we can apply the same properties of the standard NLMS algorithm [7] to our approach using the difference data set. As a consequence, the stochastic CS-LMS adaptation using an estimated gradient is convergent in the mean square if Eq. (11) is satisfied. In this case, instead of minimizing the power of the difference error in (5), the error sequence is made as smooth as possible as shown in (2).

4.1.1. Learning curves in ANC application

It is common in practice to use ensemble-average learning curves to study the statistical performance of adaptive filters. The estimation error produced by the filter in the ANC application is expressed as:

$$e(n) = s(n) + e_o(n) + \varepsilon_o^H(n)\mathbf{x}(n); \quad \text{for } \mu \text{ small} \tag{12}$$

where e_o is the estimation error produced by the Wiener filter and ε_o is the zero-order weight-error vector which satisfies the stochastic difference equation described in [4], i.e. we invoke the direct averaging method. Hence, assuming that e_o is statistically independent with $\mathbf{x}(n)$ and $s(n)$, the mean-squared error produced by the filter on the novel data is given by:

$$J(n) = J_o + E[|s(n)|^2] + E[\varepsilon_o^H(n)\mathbf{x}(n)\mathbf{x}(n)^H\varepsilon_o(n)]; \tag{13}$$

where $J_o = E[|e_o(n)|^2]$ and $J_{min} = J_o + E[|s(n)|^2]$. It can be readily seen that a reduction in mean-squared error $J(n)$ is obtained by the CS-LMS algorithm if the desired signal $s(n)$ and the input signal $\mathbf{x}(n)$ are strongly correlated since:

$$\varepsilon_o(n+1) = (I - \mu\mathbf{R}_{\mathbf{x_1}})\varepsilon_o(n) - \mu\mathbf{x_1}(n)\delta\tilde{e}_o^*(n) \tag{14}$$

where $\delta\tilde{e}_o(n) = \delta e_o(n) + \delta s(n)$ and the excess error in the steady state is expressed as:

$$\begin{aligned} J_{ex}(\infty) &= \lim_{n\to\infty} E[\varepsilon_o^H(n)\mathbf{x}(n)\mathbf{x}(n)^H\varepsilon_o(n)] = \\ &\lim_{n\to\infty} tr[\mathbf{R_x}E[\varepsilon_o(n)\varepsilon_o(n)^H]] \leq \mu\tilde{J}\sum_{k=1}^{L}\frac{\lambda_k}{2-\mu\lambda_k} < J_{ex}^{\mathrm{LMS}}(\infty) \end{aligned} \tag{15}$$

where $\tilde{J} = 2(J_{min} - Re\{r_s(1)\})$ and $r_s(1)$ is the correlation function of the desired signal at lag 1, if the input signal $\mathbf{x}(n)$ is weakly correlated (first inequality: if $R(1) \sim \mathbf{0}$) and the desired signal is strongly correlated $s(n)$ (second inequality: if $Re\{r_s(1)\} > 3/4J_{min}$), we achieve a clear reduction in excess mean squared error.

In addition, if the power of the desired signal is neglected (or disappear in time, i.e. when $s(n)$ is intermittent) the Misadjustment (M) satisfies the following equality:

$$\begin{aligned} &\mathrm{M} = \frac{J_{ex}(\infty)}{J_{min}} \simeq \mu tr(\mathbf{R}_{\mathbf{x_1}}) = \\ &\mu\left[2(tr(\mathbf{R}) - tr(Re\{\mathbf{R}(1)\}))\right] = \tfrac{1}{2}\mu_D tr(\mathbf{R}); \\ &J(n) \simeq J_o + \mu J_o tr(\mathbf{R}_{\mathbf{x_1}}) \end{aligned} \tag{16}$$

where $J_{ex}(\infty)$ is the steady state EMSE, $J_{min} = J_o$, tr stands for trace of a matrix and $\mu_D = 4\mu\left(1 - \frac{tr(Re\{\mathbf{R}(1)\})}{tr(\mathbf{R})}\right)$. In DTX systems, the adaptive filter is required to have a high step size μ to cope with changing statistics in the channel. This affects the M of the algorithm which is increased substantially. Beforehand, a clear reduction of $J_{ex}(\infty)$ in the CS-LMS algorithm is achieved by reducing J_{min} to $\tilde{J}$ (see equation 15) and/or the trace of the input correlation matrix since $\mathrm{M} \simeq \mu \cdot tr(\mathbf{R}_{\mathbf{x_1}})$ (see equations 15 and 16). The condition for the reduction of M, that depends on the input sequence, is that:

$$tr(Re\{\mathbf{R}(1)\}) > \frac{3}{4}tr(\mathbf{R}) \quad \text{or} \quad \mu_D < \mu \tag{17}$$

Note that this condition is incompatible with the one assumed to obtain the upper bound for $J_{ex}(\infty)$ ($\mathbf{R}(1) \sim \mathbf{0}$. Thus, the reduction of it by using *the trace of the correlation matrix* is unfeasible.

It also follows from the NLMS analysis [7] that the high value of μ balances the trade-off between M and the average time constant. Indeed the average time constant for the stochastic CS-LMS algorithm can be expressed as:

$$\tau \simeq \frac{L}{\mu tr(\mathbf{R}_{\mathbf{x_1}})} \tag{18}$$

where L is the filter length. On the basis of this formula, we may make the same observations as in [7] about the connection between M and τ. This constant is higher than the standard LMS algorithm for the same μ but fits with the ANC application for DTX systems (a high value of step size is needed). In the following section we show a customized version of the CS-LMS algorithm, the lag CS-LMS algorithm, that allows to select a suitable trade-off between convergence and average time constant.

5. VAD AND THE LAG CS-LMS ALGORITHM

We can exploit the advantages of the CS-LMS algorithm over noise segments just relaxing the equilibrium constraint. In many cases, constraining the least squared filter to minimize 5 is overly restrictive (see p. 171 in [8]). For example, if a delay may be tolerated, then we may consider minimizing the expected value of the difference:

$$\min_{\mathbf{w}} E[|e_k(n)|^2] \tag{19}$$

where $e_k(n) = e(n) - e(n-k)$. In most cases, a nonzero delay k will produce a better approximate filter and, in many cases, the improvement will be substantial. Following the same methodology as in section 2 and imposing the condition:

$$e^{[n+1]}(n) = e^{[n+1]}(n-k) \tag{20}$$

we obtain an additional improvement in the filtering of noise segments unlike the speech segments where the high nonstationary affects the solution provided by the lag algorithm. The explanation of this behavior is that using the novel data set $e_k(n)$ we decrease the average time constant by increasing the trace of the novel input autocorrelation matrix $\mathbf{R}_{\mathbf{x_k}}$ (and consequently increasing the M). This is really effective over noise segments as shown in Figure 2, given that a small average time constant provides a smaller averaged M over the silence frame. Thus, the combination of both operation modes using an effective VAD, i.e. based on energy and contextual information [9]-[15], is expected to supply better tracking ability and filtering results. The evaluation of the proposed strategy over an utterance of the Speech-Dat Car Aurora3 is shown in Figure 2. As clearly stated in the latter Figure and in the experimental section, the use of VAD in ANC provides an additional improvement in the results obtained by the CS-LMS algorithms separately.

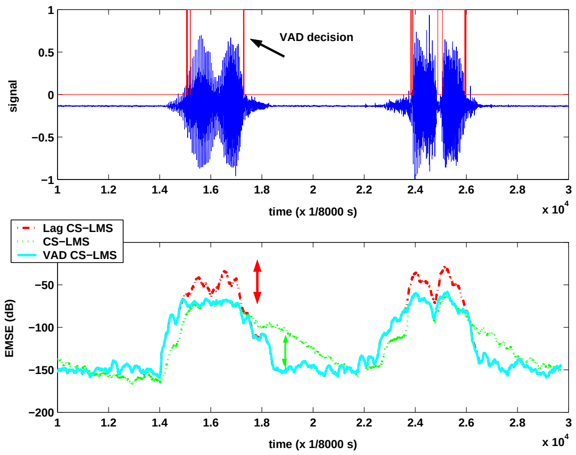

Fig. 2. The effect of combining the lag CS-LMS ($k = 5$) and CS-LMS using VAD in terms of EMSE (dB= $20 \cdot log_e(.)$).

6. EXPERIMENTS

The experimental analysis is mainly focussed on the determination of the EMSE and M at different SNR levels, step sizes, and environments. The AURORA subset of the original Spanish SpeechDat-Car database [16] was used. It contains 4914 clean recordings from more than 160 speakers. The noise ν was assumed to be zero mean white Gaussian with different variances, and non-stationary with linearly increasing variance (from $\sigma_\nu^{2,min}$ to $\sigma_\nu^{2,max}$) as shown in Table 1. The filter length L and the step size μ were also varied, $\varepsilon = 0.0001$ and the impulse response of the filters $\mathbf{h_1}$ and $\mathbf{h_2}$ were modeled as low pass IIR filters: $H_1^{-1}(z) = 1 - 0.3z^{-1} - 0.1z^{-2}; H_2^{-1}(z) = 1 - 0.2z^{-1}$. The equilibrium condition can be easily achieved over the stationary noise segments (see Figure 3 a)); over the speech segments the equilibrium constraint is hardly satisfied because of the non-stationary nature of speech. We also show the spectrogram of the estimated clean signal $e(n)$ by using the CS-LMS adaptation. Figure 3 b) shows a non-stationary experiment on the same utterance. Finally, Table 1 summarizes the *averaged results* of the EMSE and M using the proposed and referenced algorithms over the complete Spanish Aurora 3 databases, set of parameters and environments (first row: stationary environment). As a conclusion, the proposed method yields the minimum EMSE and M for a wide range of adverse noise variances, filter lengths and step sizes given in Table 1.

Table 1. Performance of Referenced and proposed LMS algorithms (dB=$20 \cdot \log_e(.)$)

Stat./Non-Stat. White Noise		N-LMS		NDN-LMS	
$\mu = 0.1; 0.25; 0.5$	$L = 8; 12; 24$	$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$	$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$
σ_ν^2	$\sigma_\nu^{2,min/max}$	-31.94	0.98	-22.83	1.67
0.01;0.1;0.5	0.1/0.5	-23,04	1,66	-11,00	1,71
VAD CS-LMS		CS-LMS		M-LMS	
$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$	$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$	$\overline{\text{EMSE}}$(dB)	$\overline{\text{M}}$
-36.69	0.38	-36.11	0.43	-32.71	1.07
-35,87	0,40	-35,31	0,45	-22,71	1,85

7. CONCLUSION

This paper showed a novel CS-LMS algorithm based on the minimization of the squared Euclidean norm of $\delta\mathbf{w}(n+1)$ subject to the constraint of equilibrium condition in the sequence of *a posteriori estimation errors* and its application to effective ANC systems. To solve this constrained optimization problem, the Lagrange multiplier method was used for the general case of complex-valued data. In this way a novel adaptation algorithm that applies non-linearities to the error and input signal sequences was obtained. In DTX systems, the trade-off between M and convergence speed was controlled i) by using a novel difference data set and ii) by including a lag in the equilibrium constraint, provided that the step size of the algorithm μ is necessarily fixed to a high value to track the variability of the incoming signal. The latter technique, which increases the trace of the input autocorrelation matrix, is only effective over noise segments since, over non-stationary speech frames, a small M is required. For both stationary and non-stationary adverse noise environments, the proposed ANC based on the CS-LMS algorithm and a VAD, which exchanges the operation modes, showed increased performance by decreasing the excess mean-squared error and misadjustment compared to referenced algorithms [3, 2, 5].

Acknowledgment

This work has been funded by PET2006-0253 and TEC2007-68030-C02-01 projects funded by the Spanish MEC, TIC-02566 of the Consejería de Innovación Ciencia y Empresa and PTEC company. This work has been written in English with the help of Mrs. M. Eugenia Cobo Lara.

8. REFERENCES

[1] B.Widrow, JR. Glover, J.M. Mccool, J. Kaunitz, C.S. Williams, R.H. Hean, J.R. Zeidler, E. Dong and R.C. Goodlin, “ANC: Principles and Applications”, *Proc. of the IEEE*, Vol. 63, No. 12, Dec. 1975.

[2] J.E. Greenberg, “Modified LMS algorithms for speech processing with an ANC”, *IEEE Trans. On Speech and Audio Processing*, vol. 6, no. 4, July 1998.

[3] S.C. Douglas and T. H.-Y. Meng, “Stochastic gradient adaptation under general error criteria”, *IEEE Trans. Signal Processing*, vol. 42, pp. 1335-1351, June 1994.

[4] H. Kushner, *Approaximation and weak convergence methods for random processes with applications to stochastic system theory*. MIT press, Cambridge, MA, 1984.

[5] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs", *IEEE Trans. Acoust., Speech, Signal Processing*. Vol ASSP-34 No4 pp 793-806. Aug 1986.

[6] S.C. Douglas, "A Family of Normalized LMS Algorithms", *IEEE Signal Processing Letters*, Vol. 1, No. 3, March, 1994

[7] S. Haykin. *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice- Hall, 1986.

[8] M. Hayes, *Statistical Digital Signal Processing and Modeling*. Ed John Wiley and Sons. 1996.

[9] J.M. Górriz, J. Ramírez and C.G. Puntonet, "An effective jointly pdf-based voice activity detector for real time applications", *IET Electronic Letters* Vol. 3 Issue 4 pp 251-253 Feb., 2007.

[10] J.M. Górriz, J. Ramírez, J.C. Segura and C.G. Puntonet. Improved MO-LRT VAD based on bispectra gaussian model. *IEE Electronics Letters*. volume 41, pp 877-879 (2005).

[11] J.M. Górriz, J. Ramírez, E.W. Lang, C.G. Puntonet. Hard C-means clustering for Voice Activity Detection. *Speech Communication Journal*. Vol 48, pp 1638-1649. (2006).

[12] J.M. Górriz, J. Ramírez, C.G. Puntonet, and J.C. Segura. Generalized LRT-based Voice Activity Detector. *IEEE Signal Processing Letters*, Vol 13 Nº 10 pp 636-639 (2006).

[13] J.M. Górriz, J.Ramírez, J.C. Segura and C.G. Puntonet. An effective cluster-based model for robust speech detection and speech recognition in noisy environments. *The Journal of the Acoustical Society of America*. vol 120,pp 470-481. 2006.

[14] J. Ramírez, J.M. Górriz, J.C. Segura, C.G. Puntonet and A. Rubio. Speech/non-speech discrimination based on contextual information integrated bispectrum LRT. *IEEE Signal Processing Letters*. pp 497-500 Issue 6, Vol 13 (2006).

[15] J. Ramírez, J.M. Górriz, J.C. Segura. Statistical voice activity detection based on integrated bispectrum likelihood ratio tests for robust speech recognition. *The Journal of Acoustical society of America*. Vol 121, Issue 5. pp. 2946-2958 (2007).

[16] A. Moreno, L. Borge, D. Christoph, R. Gael, C. Khalid, E. Stephan and A. Jeffrey. SpeechDat-Car, "A Large Speech Database for Automotive Environments", *Proceedings of the II LREC Conference*, 2000.

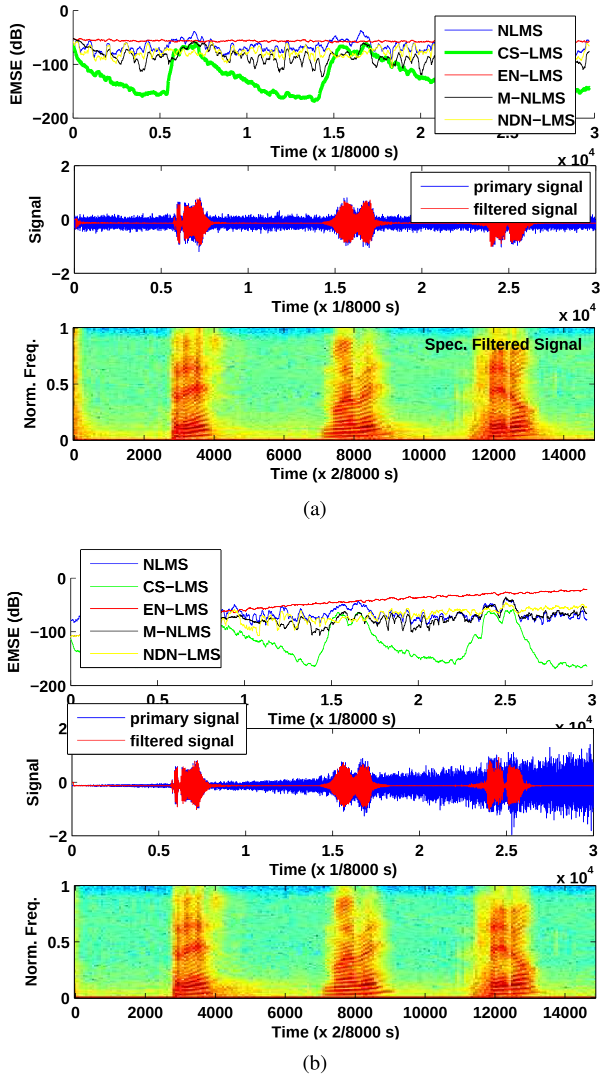

Fig. 3. EMSE evolution (dB= $20 \cdot log_e(.)$) over an utterance of the Aurora3 ($\mu = 0.1$, $L = 12$). (a) Stat. environment ($\sigma_v^2 = 0.1$). (b) Non-Stat. environment ($\sigma_v^{2,min} = 0.1$, $\sigma_v^{2,max} = 0.7$)