Speech enhancement in discontinuous transmission systems using the constrained-stability least-mean-squares algorithm

J. M. Górriz[a] and J. Ramírez
Department of Signal Theory, University of Granada, Andalucia 18071, Spain

S. Cruces-Álvarez
Department of Signal Theory, University of Seville, Seville 41004, Spain

D. Erdogmus
Dana Research Center, Northeastern University, Boston, Massachusetts 02115

C. G. Puntonet
Department of Computer Architecture and Technology, University of Granada, Andalucia 18071, Spain

E. W. Lang
Institut für Biophysik und Physikalische Biochemie, University of Regensburg, D-93040 Regensburg, Germany

(Received 16 May 2008; revised 30 September 2008; accepted 30 September 2008)

In this paper a novel constrained-stability least-mean-squares (LMS) algorithm for filtering speech sounds is proposed in the adaptive noise cancellation (ANC) problem. It is based on the minimization of the squared Euclidean norm of the weight vector change under a stability constraint over the *a posteriori* estimation errors. To this purpose, the Lagrangian methodology has been used in order to propose a nonlinear adaptation in terms of the product of differential input and error. Convergence analysis is also studied in terms of the evolution of the natural modes to the optimal Wiener–Hopf solution so that the stability performance depends *exclusively* on the adaptation parameter μ and the eigenvalues of the difference matrix $\Delta\mathbf{R}(1)$. The algorithm shows superior performance over the referenced algorithms in the ANC problem of speech discontinuous transmission systems, which are characterized by rapid transitions of the desired signal. The experimental analysis carried out on the AURORA 3 speech databases provides an extensive performance evaluation together with an exhaustive comparison to the standard LMS algorithms, i.e., the normalized LMS (NLMS), and other recently reported LMS algorithms such as the modified NLMS, the error nonlinearity LMS, or the normalized data nonlinearity LMS adaptation.
 [DOI: 10.1121/1.3003933]

PACS number(s): 43.60.Mn, 43.72.Lc, 43.72.Dv [EJS] Pages: 3669–3683

I. INTRODUCTION

The widely used least-mean-squares (LMS) algorithm has been successfully applied to many filtering applications including modeling,[1] equalization, control,[2,3] echo cancellation biomedicine, or beamforming.[4] One of these important applications is found in voice communication systems such as dialogue systems in airplane or helicopter cockpits where the extraction of noise from the desired speech is required. Another major application concerns filtering biosignals in adverse noise conditions that usually are present in medical instruments. These applications include the cancellation of several kinds of interference in electrocardiograms (ECGs) such as canceling the power line interference, the donor ECG in heart-transplant electrocardiography, or the maternal ECG in fetal electrocardiography;[5] and adaptive noise canceling (ANC) in modern techniques such as positron emission tomography or single positron emission computed tomography images.[6]

The typical noise cancellation scheme of a voice communication system is shown in Fig. 1. Two distant microphones are needed for such application capturing 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. Note that the primary signal $d(n)$ consists of the clean speech $s(n)$ and the additive noise $v_1(n)$ recorded by the primary microphone. The reference noise signal $v_2(n)$ is the input of the FIR filter defined by the $L\times 1$ weight vector $\mathbf{w}(n)$. Its output $y(n)$ is the "best" estimation of the additive noise in the primary signal. The criterion for the best estimation is the minimization of the expected value of the squared sequence $e(n)$, which tends to equal the original signal $s(n)$. The weights are typically adjusted using the LMS algorithm[5] because of its simplicity, ease of implementation, and low computational complexity. The weight update equation for ANC is

[a]URL: http://www.ugr.es/~gorriz. Electronic mail: gorriz@ugr.es

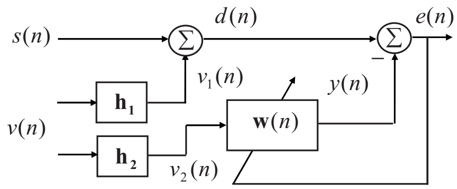

FIG. 1. Adaptive noise canceler.

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

where μ is a step-size parameter with units of inverse power, $e(n) = d(n) - y(n) = s(n) + v_1(n) - \mathbf{w}^H(n)\mathbf{x}(n)$ is the system output, where ^{H} stands for conjugate transpose, $*$ denotes complex conjugate, and $\mathbf{x}(n) = (x(n), \ldots, x(n-L-1))^T$ is the data vector containing L samples of the reference signal, $v_2(n)$, that are present in the adaptive filter's tapped delay line at time n. Many ANCs (Refs. 5 and 7–10) have been proposed in the past years using modified LMS (MLMS) algorithms in order to simultaneously improve the tracking ability and speed of convergence. In this way, many efforts have been directed to reduce computational load and power consumption in adaptive filter implementations.[11,12]

On the other hand, the emerging applications of speech technologies (particularly in mobile communications, robust speech recognition, or digital hearing aid devices) often require a noise reduction scheme working in combination with a precise voice activity detector (VAD).[13,14] During the past decade numerous researchers have studied different strategies for detecting speech in noise and the influence of the VAD decision on speech processing systems.[15–21] The nonspeech detection algorithm is an important and sensitive part of most of the existing single-microphone noise reduction schemes.[22,23] The VAD is even more critical for nonstationary noise environments since it is needed to update the constantly varying noise statistics. A VAD achieves silence compression in modern mobile telecommunication systems reducing the average bit rate by using the discontinuous transmission (DTX) mode.[16] VADs have been used extensively in the open literature as a reference for assessing the performance of new algorithms.

This paper shows a novel adaptation in combination with a VAD to enhance speech signals in DTX systems, which are characterized by a sudden change in channel signal statistics. The algorithm is derived assuming stability in the sequence of *a posteriori* errors instead of the more restrictive hypothesis used in previous approaches,[24] that is, enforcing it to vanish. The result of the Lagrange minimization is the application of the normalized LMS (NLMS) algorithm to a new set that consists of difference signals that are more suitable for ANC of speech signals in DTX systems. In this way, the sudden transitions in the channel are counteracted by the use of the concurrent change of the processing variables. The deterministic version of the proposed adaptation is also shown to be equivalent to the gradient descent method, from which the NLMS algorithm derives, in the sense that they have the same optimal solution.

The paper is organized as follows. In Sec. I A the background on LMS algorithms applied to the ANC problem is summarized. Then a novel constrained-stability (CS) LMS algorithm for filtering speech signals in additive noise is proposed in Sec. II. This is achieved by minimizing the norm of the difference weight vector and imposing a stability constraint in the sequence of errors in order to reach the *equilibrium condition*. In Sec. III A a connection to the previous LMS algorithm is shown, which allows to analyze: the conditions for convergence. This novel methodology for weight adaptation provides both higher filtering performance as well as enhanced tracking ability over the traditional LMS methods in context of ANC applications. In addition, an improvement in filtering results by means of a VAD is also achieved in Sec. IV. The VAD scheme combines the operation of the proposed CS-LMS with a suitable MLMS algorithm, the lag CS-LMS, which is intended for filtering nonstationary noise segments, exclusively. Thus, we select the best adaptation on each kind of frame (speech or noise) achieving an additional improvement in filtering performance and a clear reduction of background noise, on the filtering algorithms operating separately. The experimental analysis conducted in Sec. V provides the information on how the proposed and referenced algorithms operate on discontinuous speech segments in the ANC application. Finally, we state some conclusions in Sec. VI.

A. Background on modified LMS algorithms

Many researchers have proposed modifications to the LMS algorithm, which are motivated by two factors: (i) reducing the computational burden, and (ii) improving the trade-off between convergence time and steady-state performance. One approach for improving LMS algorithms is to modify the time-varying step size, i.e., applying nonlinearities to the quantities used to update the weights (the data vector and/or the output signal). For instance, Bershad studied the performance of the NLMS with an adaptive step size $(\mu_{adap}) = \mu / \|\mathbf{x}(n)\|^2$ in Ref. 25, showing advantages in convergence time and steady state. Later, Douglas and Meng[9] proposed the optimum nonlinearity for any input probability density of the independent input data samples, obtaining the normalized data nonlinearity LMS (NDN-LMS) adaptation

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{1 + \mu\|\mathbf{x}(n)\|^2} e^*(n)\mathbf{x}(n). \quad (2)$$

Although this algorithm is designed to improve steady-state performance, its derivation did not consider the ANC with a strong target signal in the primary input.[7] Greenberg's MLMS 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 unknown *a priori* information of the processes

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu_L}{(e^2_{min}(n) + \mu_L\|\mathbf{x}(n)\|^2)} e^*(n)\mathbf{x}(n), \quad (3)$$

where $\mu_L = \mu_{opt}/L$ and μ_{opt} is a dimensionless step-size parameter. Further e_{min} is the Wiener–Hopf solution in stationary environment. Recently, an interesting approach has been proposed based on a nonlinearity applied to the data vector exclusively.[10] The error nonlinearity LMS (EN-LMS) algorithm is described in terms of the following LMS adaptation:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\alpha_{\text{opt}}}{e + \|\mathbf{e}(n)\|^2} e^*(n)\mathbf{x}(n), \qquad (4)$$

where $\|\mathbf{e}(n)\|^2 = \sum_{i=0}^{n-1} |e(n-i)|^2$ is the squared norm of the error vector, which is estimated over its entire history. As in the NLMS algorithm, ϵ is usually added to prevent instability in the variable step size of the algorithm if $\|\cdot\|^2$ is too small. For both stationary and nonstationary noise environments, the EN-LMS algorithm has shown a good trade-off between speed of convergence and misadjustment (M) (Ref. 26) improving the performance over Douglas' NDN-LMS.[10]

II. CONSTRAINED-STABILITY (CS)-LMS ALGORITHM

In this section, a novel adaptation that applies nonlinearities to the error sequence and the input signal in terms of difference values is presented. The main motivation of the CS-LMS algorithm is to obtain a suitable adaptation for DTX systems, in which the sudden transitions in the channel, due to the presence of intermittent signals, affect the performance of the standard filtering algorithms because of their limited tracking ability. The adaptation is derived as a solution to a constrained optimization problem,[27] by using the Lagrangian formulation that minimizes the norm of the difference weight vector under the *stability constraint*, which makes the error sequence as constant as possible. Our approach belongs to the class of MLMS algorithms based on applying nonlinearities to one or both of the quantities used to update the adaptive weights (the data vector and/or the output signal),[9,24,28,29] rather than optimizing the step-size function.[7,30] The derivation of the nonlinearity applied to the *target signal and the data vector* is reached as a generalization of the NLMS algorithm, which may be also obtained using this formulation indeed.[27] In this case the constraint of the standard LMS algorithm (the error sequence must be canceled) is relaxed to be as smooth as possible. In this way, consider the following constrained optimization problem:

$$\text{minimize} \|\delta\mathbf{w}(n+1)\|^2$$

$$\text{subject to the constraint } e^{[n+1]} = e^{[n+1]}(n-1), \qquad (5)$$

where $\delta\mathbf{w}(n+1) = \mathbf{w}(n+1) - \mathbf{w}(n)$, and $e^{[k]}(n) = d(n) - \mathbf{w}^H(k)\mathbf{x}(n)$ is the error at time n using the weight vector at time k. The method of Lagrange multipliers converts the problem of constrained minimization into one unconstrained minimization by introducing Lagrange multipliers

$$\mathcal{L}(\mathbf{w}(n+1)) = \|\delta\mathbf{w}(n+1)\|^2 + \text{Re}[\lambda^*(e^{[n+1]}(n+1) - e^{[n+1]}(n))]. \qquad (6)$$

The solution to this optimization problem $\mathbf{w}^{\text{opt}}(n+1)$ minimizes the norm of the difference between two consecutive weight vectors and satisfies the equilibrium constraint or condition $e^{[n+1]}(n) = e^{[n+1]}(n-1)$. This equilibrium constraint imposes *stability* on the sequence of *a posteriori errors*, i.e., the optimal solution $\mathbf{w}^{\text{opt}}(n+1)$ that is obtained from the minimization of Eq. (5) under the constraint is the one that renders the sequence of errors as smooth as possible. Taking partial derivative in Eq. (6) with respect to $\mathbf{w}^H(n+1)$ and setting it equal to zero, the Lagrange multiplier is found to be (the complete derivation can be seen for clarity in Appendix A)

$$\lambda = \frac{2\,\delta e^{[n]}(n)}{\|\delta\mathbf{w}(n)\|^2}, \qquad (7)$$

where $\delta e^{[n]}(n) = e^{[n]}(n) - e^{[n]}(n-1)$ is the difference of 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 of consecutive input vectors. Thus, the minimum of the Lagrangian function satisfies the following adaptation equation:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\|\delta\mathbf{w}(n)\|^2 + \varepsilon} \delta\mathbf{x}(n)(\delta e^{[n]}(n))^*. \qquad (8)$$

This algorithm for adapting the weight vector is based on difference quantities and the concept of the equilibrium condition given by the constraint. It is interesting to notice that the equilibrium condition enforces the convergence of the algorithm if $\|\delta\mathbf{x}(n)\|^2 \neq 0$. The adaptation has been completed by introducing ϵ to prevent instability for small $\|\delta\mathbf{x}(n)\|^2$ and the constant step μ to control the speed of the adaptation. It can be seen readily that the computational load of the proposed adaptation is of the same order as with the standard LMS algorithm. The CS-LMS algorithm requires only $2L+1$ complex multiplications and $3L+1$ complex additions per iteration, where L is the number of tap weights used in the adaptive filter. Thus, the computational complexity of the algorithm is $\mathcal{O}(L)$. In addition, like the NLMS algorithm, the CS-LMS adaptation requires computation of the normalization term $\|\delta\mathbf{x}(n)\|^2$, which involves only two squaring operations, one addition and one substraction if it is evaluated recursively.

As shown in following sections, 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[26] since the adaptation rule is *equivalent* except for the difference data. Then, we will show that the EMSE of the proposed adaptation algorithm is lower than the one obtained with the standard NLMS algorithm in DTX systems, i.e., an intermittent, strongly correlated desired signal under an adverse noisy environment. 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 $\mathbf{w}_0$.

III. THE OPTIMAL SOLUTION OF THE CS-LMS ADAPTATION

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 Wiener solution in the following.

Theorem 1 (convergence equivalence): *Consider a transversal filter with tap input in vector* $\mathbf{x}(n)$ *and a corresponding set of tap weights in vector* $\mathbf{w}(n)$. *By comparing the estimate given by the filter* $y(n)$ *with the desired response*

$d(n)$ we produce the estimation error denoted by $e(n) = d(n) - y(n)$. If the desired signal $d(n)$ is generated by the multiple linear regression model, i.e., $d(n) = \mathbf{w}_0^H \mathbf{x}(n) + e_0(n)$, where $e_0(n)$ is an uncorrelated white-noise process that is statistically independent of the input vector $\mathbf{x}(n)$, then the optimal solution of the deterministic CS-LMS algorithm is the Wiener solution $\mathbf{w}_0$.

Proof: See Appendix B.

Thus, the Wiener solution is the same for both LMS and CS-LMS algorithms and then the deterministic CS-LMS algorithm is justified. As a consequence, the stochastic CS-LMS adaptation using an estimated gradient, or equivalently the NLMS algorithm over the difference data set, will converge in the mean to the Wiener solution. In this case, instead of minimizing the power of the difference error in Eq. (B1) we try to make the error sequence as smooth as possible as shown in Eq. (6).

A. A connection between CS-LMS and LMS adaptations

From Eq. (8) we can derive an interesting connection between our adaptation and the classical LMS algorithm. The classical LMS adaptation is obtained using the simplest way to estimate the deterministic gradient vector $\nabla J(n)$ in the steepest descent (SD) algorithm. In this way, we use instantaneous values $\hat{\nabla} J(n) = -e(n)\mathbf{x}(n)$, where $J(n) = E[|e(n)|^2]$ is the mean squared error (MSE) function. Expressing Eq. (8) in terms of the differences of the error sequence $e(n)$ and the input vector $\mathbf{x}(n)$ and redefining the variable step size as $\hat{\mu}$, we find that

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \hat{\mu}[(e^{[n]}(n))^*\mathbf{x}(n) + (e^{[n]}(n-1))^*\mathbf{x}(n-1) - (e^{[n]}(n))^*\mathbf{x}(n-1) - (e^{[n]}(n-1))^*\mathbf{x}(n)]. \tag{9}$$

The first two elements in the adaptation correspond to the instantaneous estimation of the gradient vector $\nabla J^{[n]}(k) = \nabla E[|e^{[n]}(k)|^2]$ at times $k = n$ and $n-1$, in the direction of minimization. The second two terms correspond to the instantaneous estimation of the gradient vectors $\nabla \hat{J}^{[n]}(n) = \nabla E[e^{[n]}(n)(e^{[n]}(n-1))^*] = \nabla r_e(1)$ and $\nabla \tilde{J}^{[n]}(n-1) = \nabla E[e^{[n]}(n-1)(e^{[n]}(n))^*] = \nabla r_e(-1)$, where $r_e(k)$ denotes the correlation function of the error sequence and $\tilde{J}^{[n]}(n-1) = (\hat{J}^{[n]}(n))^*$, in the direction of maximization [we assume $e(n)$ is wide sense stationary then the correlation function is symmetric for real variables: $r_e(k) = r_e(-k)$]. Thus, our algorithm can be viewed as a pair of LMS or SD adaptations at times $n, n-1$ and a pair of adaptations that try to *maximize* the correlation function at lag 1. The maximum value of the correlation function is achieved for lag $k=0$; thus, the second pair of adaptations tries to equalize $e^{[n]}(n) = e^{[n]}(n-1)$ in order to render $r_e(1)$ as close as possible to $r_e(0)$. This is equivalent to the above mentioned *equilibrium condition* that we imposed in the Lagrange formulation in order to solve the filtering problem.

B. Relationship between stochastic information gradients and CS-LMS

Several learning algorithms, where the learning relies on the concurrent change of processing variables, have been proposed in the past for decorrelation, blind source separation, or deconvolution applications.[31,32] Classical approaches propose a cost function estimator and a gradient descent learning algorithm based adaptation to find the optimal solution. The stochastic information gradient (SIG) algorithms[32] maximize (or minimize) Shannon's entropy of the sequence of error terms using an instantaneous value based estimator of the probability density function (pdf) and Parzen windowing. Given a random variable Y with pdf $f_Y(y)$ Shannon's entropy and its stochastic entropy estimator are expressed as

$$H_S(Y) = -\int_{-\infty}^{\infty} f_Y(y)\log(f_Y(y))dy = E_Y[-\log(f_Y(y))] \approx \hat{H}_S(Y) = -\log(f_Y(y(k))), \tag{10}$$

where $y(k)$ denotes the most recent sample at time k. Since in practice the pdf of Y is unknown a biased Parzen window estimate is utilized as follows:

$$\hat{f}_Y(y(k)) = 1/N \sum_{i=k-N}^{k-1} \kappa_\sigma(y(k) - y(i)), \tag{11}$$

where N is the window length and $\kappa_\sigma(x)$ is the kernel function with size σ.[32] The result for the maximization (or minimization) of Eq. (10) using Eq. (11) with respect to a set of parameters $\mathbf{w}$, which generates $Y = g(\mathbf{w})$, is an adaptation depending on fixed-width σ kernel functions.[32] If we select $N = 1$ and Gaussian kernels the adaptation reduces to

$$\delta\mathbf{w}(n+1) = \frac{\mu}{\sigma^2}(y(k) - y(k-1))\left(\frac{\partial y(k)}{\partial \mathbf{w}} - \frac{\partial y(k-1)}{\partial \mathbf{w}}\right). \tag{12}$$

The latter adaptation is essentially the CS-LMS algorithm except for the constant factor given the fixed kernel size (the maximization of Shannon's entropy is related, in some sense, to the description made in the above section in terms of correlation functions). Our approach can be described as the maximization of Shannon's entropy estimator but using a variable-size σ kernel, which experimentally provides a better estimator for the unknown pdf. The benefits of using variable kernel density estimators are clear from works such as Refs. 33–35. We refer the interested reader to them for a treatment of the problem of kernel size estimation, which is beyond the scope of this paper.

The CS-LMS and SIG methodologies can be further understood in the greater context of errors-in-variables techniques; specifically in the field of adaptive system identification also referred to as instrumental variable approach. The connection between CS-LMS presented in Eqs. (8), (B1), and (B2) and the error whitening criterion proposed by Rao *et al.*[36] should be noted at this point. Utilizing difference input and error data effectively corresponds to exploiting autocorrelation statistics at lags other than zero. The benefits of such departures from classical error-energy (zero-lag error autocorrelation) measures of performance enables the filters to become more robust to noise present in the signals.

IV. CONVERGENCE ANALYSIS OF THE CS-LMS ALGORITHM

Let us first consider the convergence of the deterministic CS-LMS algorithm, which is described as the SD algorithm on the novel data set. Taking the connection expressed in Eq. (9) into consideration we can formulate the following deterministic adaptation for the CS-LMS:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \hat{\mu}[-\nabla\{J^{[n]}(n) + J^{[n]}(n-1)\} + 2\,\mathrm{Re}\{\nabla \hat{J}^{[n]}(n)\}]. \tag{13}$$

In terms of the novel data model $\{\mathbf{x}_1(n), d_1(n), e_1(n)\}$ introduced in Appendix B the adaptation could be expressed as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \hat{\mu}[-\nabla\{J_1^{[n]}(n)\}], \tag{14}$$

where $J_1^{[n]}(n) = E[|e_1(n)|^2]$. Since $e_1(n) = d_1(n) - \mathbf{w}^H\mathbf{x}_1(n)$ then

$$J_1^{[n]}(n)\sigma_{d_1}^2 - \mathbf{w}^H\mathbf{r}_{\mathbf{d_1x_1}}\mathbf{r}_{\mathbf{d_1x_1}}{}^H\mathbf{w} - \mathbf{w}^H\mathbf{R}_{\mathbf{x_1}}\mathbf{w}. \tag{15}$$

As shown in Appendix C, a necessary and sufficient condition for the convergence can be expressed, in terms of the step-size parameter $\hat{\mu}$ and the largest eigenvalue $\lambda_{\max}$ 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

$$0 < \hat{\mu} < \frac{1}{\lambda_{\max}}. \tag{16}$$

A. Convergence analysis of the stochastic CS-LMS algorithm

From Eq. (8) it can be seen readily 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 in Sec. III that (i) they have the same Wiener solution and (ii) the stochastic CS-LMS is equivalent to the NLMS on the novel data set, we can assume the same properties of the standard NLMS algorithm[26] to our approach using the difference data set. As a consequence, the stochastic CS-LMS adaptation using an estimated gradient is convergent in the MSE sense if Eq. (16) is satisfied. In this case, instead of minimizing the power of the difference error in Eq. (B1) the error sequence is made as smooth as possible as shown in Eq. (6). Further the evolution of the expected value of the stochastic weight vector $\mathbf{w}(n)$ also satisfies Eq. (C3) as in the deterministic CS-LMS algorithm.

1. Learning curves in the ANC application

It is common in practice to use ensemble-average learning curves to study the statistical performance of adaptive filters. In the ANC application the derivation of these curves is slightly different due to the presence of the desired clean signal $s(n)$. The estimation error produced by the filter in the ANC application is expressed as

$$e(n) = s(n) + e_0(n) + \varepsilon_0^H(n)\mathbf{x}(n) \quad \text{for } \mu \text{ small}, \tag{17}$$

where e_0 is the estimation error produced by the Wiener filter and ε_0 is the zero-order weight-error vector, which satisfies the stochastic difference equation described in Ref. 37, i.e., we invoke the direct-averaging method. Hence, assuming that e_0 is statistically independent of $\mathbf{x}(n)$ and $s(n)$, the MSE produced by the filter on the novel data is given by

$$J(n) = J_0 + E[|s(n)|^2] + E[\varepsilon_0^H(n)\mathbf{x}(n)\mathbf{x}(n)^H\varepsilon_0(n)], \tag{18}$$

where $J_0 = E[|e_0(n)|^2]$ and $J_{\min} = J_0 + E[|s(n)|^2]$. It can be seen readily that a reduction in MSE $J(n)$ is obtained by the CS-LMS algorithm if the desired signal $s(n)$ and the input signal $\mathbf{x}(n)$ are either strongly and weakly correlated, respectively (see Appendix D). Since the evolution of the weight-error vector is

$$\varepsilon_0(n+1) = (I - \mu\mathbf{R}_{\mathbf{x_1}})\varepsilon_0(n) - \mu\mathbf{x_1}(n)\,\delta\tilde{e}_0^*(n), \tag{19}$$

where $\delta\tilde{e}_0(n) = \delta e_0(n) + s_1(n)$ and the excess error in the steady state is expressed and bounded as

$$\begin{aligned} J_{\mathrm{ex}}(\infty) &= \lim_{n\to\infty} E[\varepsilon_0^H(n)\mathbf{x}(n)\mathbf{x}(n)^H\varepsilon_0(n)] \\ &= \lim_{n\to\infty} \mathrm{tr}[\mathbf{R_x}E[\varepsilon_0(n)\varepsilon_0(n)^H]] \\ &\le \mu\tilde{J}\sum_{k=1}^{L}\frac{\lambda_k}{2-\mu\lambda_k} < J_{\mathrm{ex}}^{\mathrm{LMS}}(\infty), \end{aligned} \tag{20}$$

where $\tilde{J} = 2(J_{\min} - \mathrm{Re}\{r_s(1)\})$ 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 $\mathbf{R}(1) \sim \mathbf{0}$] and the desired signal is strongly correlated $s(n)$ (second inequality: if $\mathrm{Re}\{r_s(1)\} > 3/4 J_{\min}$, we achieve a clear reduction in excess MSE.

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

$$\begin{aligned} M &= \frac{J_{\mathrm{ex}}(\infty)}{J_{\min}} \simeq \mu\,\mathrm{tr}(\mathbf{R}_{\mathbf{x_1}}) = \mu[2(\mathrm{tr}(\mathbf{R}) - \mathrm{tr}(\mathrm{Re}\{\mathbf{R}(1)\}))] \\ &= \tfrac{1}{2}\mu_D\,\mathrm{tr}(\mathbf{R}), \end{aligned}$$

$$J(n) \simeq J_0 + \mu J_0\,\mathrm{tr}(\mathbf{R}_{\mathbf{x_1}}), \tag{21}$$

where $J_{\mathrm{ex}}(\infty)$ denotes the steady-state EMSE, $J_{\min} = J_0$, tr stands for the trace of a matrix, and $\mu_D = 4\mu\,(1 - \mathrm{tr}(\mathrm{Re}\{\mathbf{R}(1)\})/\mathrm{tr}(\mathbf{R}))$.

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 misadjustment (M) of the algorithm, which is increased substantially. Beforehand, a clear reduction of $J_{\mathrm{ex}}(\infty)$ of the CS-LMS algorithm is achieved by reducing $J_{\min}$ to $\tilde{J}$ [see Eq. (20)] and/or the trace of the input correlation matrix since $M \simeq \mu\,\mathrm{tr}(\mathbf{R}_{\mathbf{x}_1})$ [see Eqs. (20) and (21)]. The condition for the reduction of M, which depends on the input sequence, is the following:

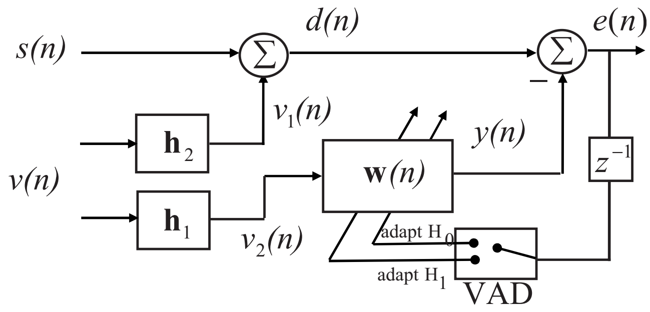

FIG. 2. Improving adaptive noise canceler using VAD.

$$\mathrm{tr}(\mathrm{Re}\{\mathbf{R}(1)\}) > \tfrac{3}{4}\mathrm{tr}(\mathbf{R}) \quad \text{or } \mu_D < \mu. \tag{22}$$

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

It also follows from the NLMS analysis[26] 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

$$\tau \simeq \frac{L}{\mu\,\mathrm{tr}(\mathbf{R}_{\mathbf{x}1})}, \tag{23}$$

where L is the filter length. On the basis of this formula, we may make the same observations as in Ref. 26 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 DXT 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 over noise segments.

V. IMPROVING SPEECH FILTERING IN ANC USING VAD

Based on the properties and production of the speech signal, different assumptions can be derived to allow distinction between speech periods and noise-only periods. The equilibrium condition in Eq. (5) is hardly met in distinctly nonstationary segments [of course the condition imposed by the NLMS algorithm is even more unfeasible, $e^{[n+1]}(n)=0$, see Ref. 26], i.e., in DTX systems, these segments correspond to the presence of noisy speech ($t\sim 0.2$ s). However, background noise is usually assumed stationary for much longer periods, e.g., $t>1$ s, than speech. Thus, a better performance of the algorithm is expected in noise segments in which the equilibrium condition could be even more *relaxed*. In the past, the use of a VAD has been critical in attaining a high level of performance in speech processing systems, even more, for nonstationary noise environments since it is needed to update the constantly varying noise statistics.[38] In this section we propose the use of a standard energy-based VAD in order to exchange the application of different filtering algorithms providing a better tracking of the variability of statistics in noise/speech transitions as shown in Fig. 2. One solution is to apply CS-LMS to noise segments (H_0), where the equilibrium condition can be easily achieved, and a MLMS algorithm,[7] using the initial weight vector provided by the CS-LMS, on speech segments (H_1) (see Fig. 3). As shown in Sec. VI, the combination of both methods provides synergy that would be expected to enhance the filter effectiveness over the referenced algorithms. Based on the same philosophy, we propose the lag-CS-LMS algorithm that relaxes the equilibrium constraint over noise segments, and combined with the CS-LMS over speech segments, obtains a better performance in terms of EMSE and M than all the referenced filtering algorithms.

A. A filtered energy-based VAD

In order to avoid the inconvenience of high noise level conditions in VAD,[15] we use the filtered sequence of *a priori* errors $e^{[n]}(n)$ of the ANC to detect the presence of speech segments. The use of the reference signal $d(n)$ for this task would cause a poor hit rate of speech/nonspeech segments at high noise conditions. Then the application of VAD to this speech processing system would be completely unnecessary, i.e., it would show a low level of performance.

Short time average energy is the most evident way to classify signal to speech and nonspeech periods, since the signal is assumed to have higher energy, when speech is present in the filtered sequence. The short time average energy-based decision at time $n+1$ can be described as

$$E(n+1) = \frac{1}{N}\sum_{k=0}^{N-1}(e^{[n]}(n-k))^2 \underset{H_0}{\overset{H_1}{\gtrless}} \eta(n), \tag{24}$$

where N denotes the amount of contextual information included in the VAD decision and $\eta(n)$ is an empirical threshold that models the background noise. In particular, we compute the average energy of a set of initial noise segments and update it according to the VAD decision (during nonspeech periods), that is, in order to adapt the operation to nonstationary and noisy environment we compute

$$n(n+1) = \alpha\eta(n) + (1-\alpha)E(n), \tag{25}$$

where $\alpha=0.9$ for a soft decision function. To overcome the difficulties in the low signal-to-noise ratio (SNR) regime subband energies have been employed.[18–20] Subband energies will describe the energy distribution of the signal in frequency domain with some predefined resolution. Energy-based VAD algorithms can provide a good performance, when the energies of the speech periods are significantly higher than the energies of the background noise-only periods. Most of the features used in VAD algorithms are related to energy[15–20] indeed. However, when the energy of the background noise is comparable, the signal energy alone will result in a poor performance. Moreover, the unvoiced parts of the speech will have low energy constantly, which makes the classification of the unvoiced phonemes more difficult than the ones with voice present. Nevertheless, this situation can be avoided using the filtered sequence provided by the ANC that is characterized by a high SNR in the steady state of the filter. On one hand, this allows the use of effective classical VADs, i.e., based on energy to exchange the operation of the ANC; on the other hand, it provides a small delay

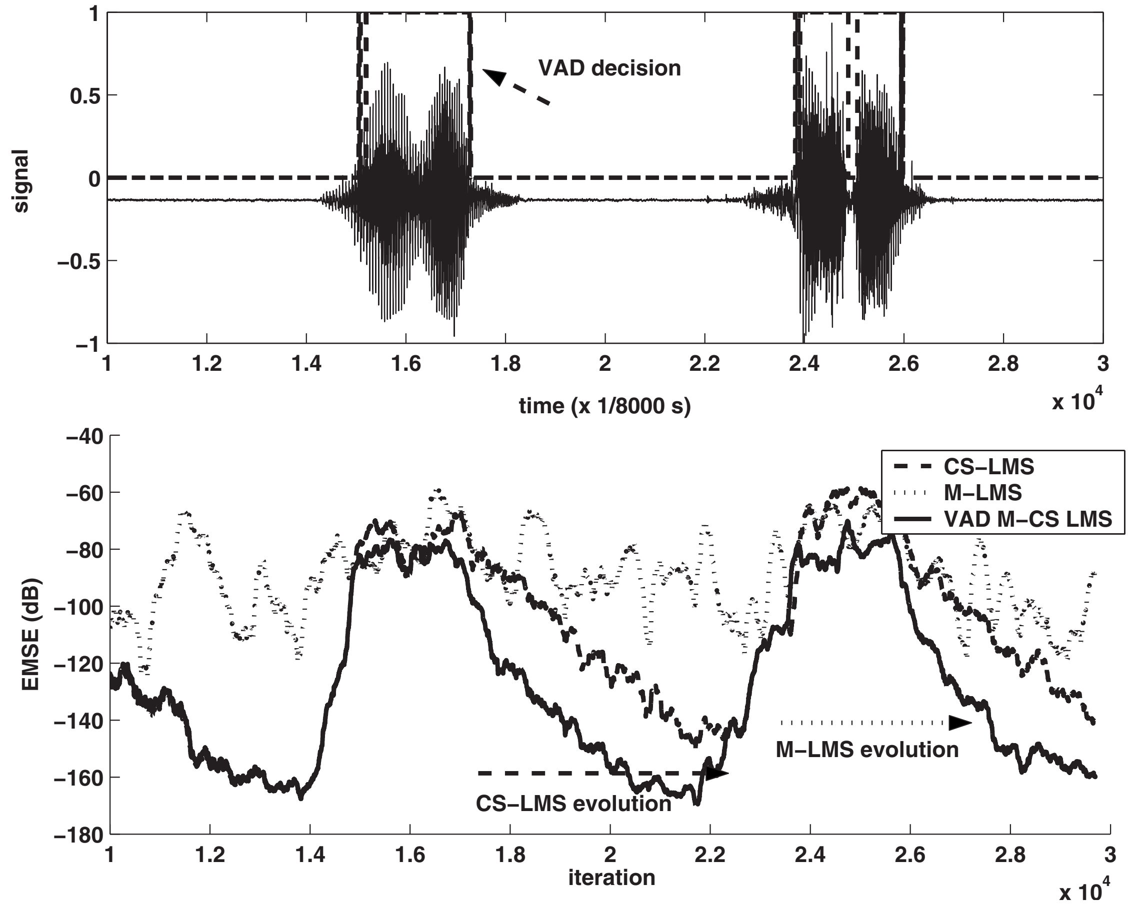

FIG. 3. ANC application: The effect of combining MLMS and CS-LMS over an utterance of AURORA 3 database using VAD in terms of EMSE [EMSE in dB=20 $\log_e(.)$; Gaussian noise with variance $\sigma_v^2=0.01$, $L=12$, $\mu=0.1$, and $\mathbf{h}_1$, $\mathbf{h}_2$ as defined in Sec. VI].

to the VAD that detects the presence of speech at a later stage. Regardless, that is not a serious implementation obstacle since the ANC works sample by sample unlike most VADs that works over contextual information windows.[18]

B. The lag-CS-LMS algorithm

We can exploit the advantages of the CS-LMS algorithm in noise segments just relaxing the equilibrium condition. In many cases, constraining the least squares filter to minimize Eq. (B1) is overly restrictive (see page 171 in Ref. 39). For example, if a delay may be tolerated, then we may consider minimizing the expected value of the difference as follows:

$$\min_{\mathbf{w}} E[|e_k(n)|^2], \tag{26}$$

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 Sec. II and imposing the condition

$$e^{[n+1]}(n) = e^{[n+1]}(n-k), \tag{27}$$

we obtain an additional improvement in the filtering of noise segments unlike the speech segments where the high nonstationary noise affects the solution provided by the algorithm. The explanation of this behavior is that using the novel data set $e_k(n)=e(n)-e(n-k)$, which is increasing the lag k, we decrease the average time constant τ in Eq. (C10) by increasing the trace of the input autocorrelation matrix $\mathbf{R}_{\mathbf{x}_k}$ (and consequently increasing M). This is really effective over noise segments as shown in Fig. 4, given that a small average time constant provides a smaller averaged M over nonspeech frames. Thus, the combination of both operation modes using an effective VAD, i.e., based on energy and contextual information,[20] is expected to supply better filtering performance. The evaluation of the proposed strategy over an utterance of the Speech-Dat-Car (SDC) AURORA 3 is shown in Fig. 4. As clearly stated in the latter figure and in Sec. VI, the use of VAD in ANC provides the best results obtained by the CS-LMS algorithms separately. The selection of lag k is motivated by the trade-off between the M and the average time constant (see Fig. 5). In this case, given a large step size $\mu=0.25$, a higher convergence speed does not reduce the averaged EMSE in noise segments as clearly stated in Fig. 5(a). However, when the specifications of the problem require a good filtering performance (i.e., a reduced M), a smaller step size $\mu=0.1$ should be selected. Then the averaged EMSE in noise segments may be reduced by incorporating a lag into the algorithm, which speeds up the algorithm convergence (see also the experimental analysis in Sec. VI). In Fig. 5 the robustness of the CS-LMS algorithm against the noise power is also highlighted.

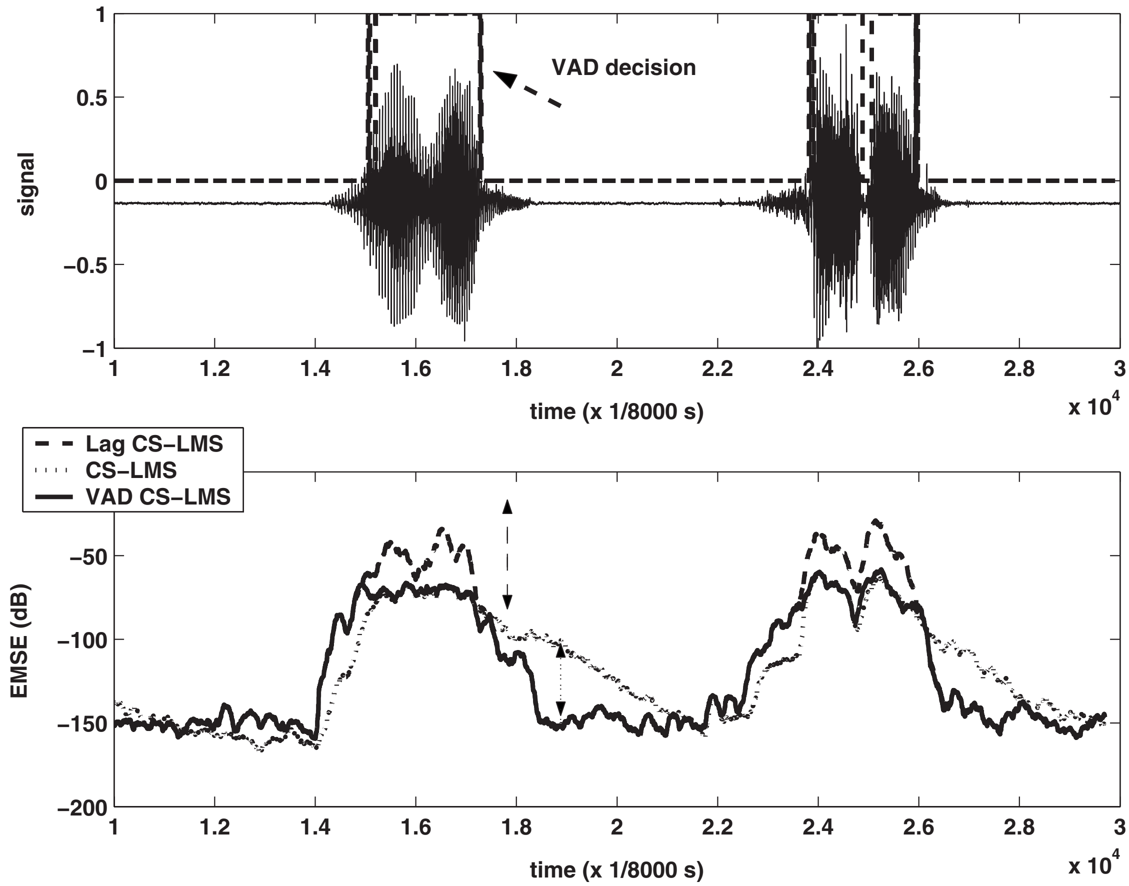

FIG. 4. ANC application: The effect of combining the lag CS-LMS ($k=5$) and CS-LMS using VAD in terms of EMSE. [EMSE in dB$=20\log_e(.)$; Gaussian noise with variance $\sigma_v^2=0.01$, $L=12$, $\mu=0.1$, and $\mathbf{h}_1$, $\mathbf{h}_2$ as defined in Sec. VI].

VI. EXPERIMENTS

Several experiments are commonly conducted in order to evaluate the performance of the CS-LMS algorithm. The experimental analysis is mainly focused on the determination of the EMSE and misadjustment (M) (Ref. 26) at different SNR levels, step sizes, and environments. This section describes the experimental framework and the objective performance measures conducted to evaluate the proposed algorithm. For this purpose, the AURORA subset of the original Spanish SDC database[40] was used. This database contains 4914 recordings using close-talking (CH0) and distant hands-free microphone (CH1) from more than 160 speakers. The files are categorized into three noisy conditions: quiet, low, and high noise conditions, which represent different driving conditions with average SNR values between 25 and 5 dB. The EMSE and M were determined for each recording using the close-talking microphone high noise condition for the CS-LMS, NDN-LMS,[9] NLMS,[25] EN-LMS,[10] and MLMS (Ref. 7) algorithms. The *experimental* EMSE at the kth iteration is defined by

$$\text{EMSE}(k) = \frac{1}{j}\sum_{j=1}^{J} |e(k-j) - s(k-j)|^2, \qquad (28)$$

where J is the number of samples used in the estimation of the EMSE. On the other hand, M is defined as a normalized MSE, i.e., the ratio of the steady-state EMSE ($J_{\text{ex}}(\infty)$) to the minimum MSE ($J_{\min}$). The MSE is obtained removing $s(n)$ in Eq. (28) [estimation of the power of $s(n)$].

A. Numerical experiment

The initial set of simulations used a simple ANC configuration to test the analytical results derived in Sec. IV. In this case, the desired signal $s(n)$ is a sum of an intermittent zero-mean AR(1) process with variance 1 and its pole at $a_s(1)=0.99$ and a zero-mean additive white noise with variance 0.001. The AR(1) process turns on and off every 3000 samples. The noise source $v(n)$ is a zero-mean Gaussian process with variance 1, and it is assumed to be independent of $s(n)$. The impulse response of the filters $\mathbf{h}_1$ and $\mathbf{h}_2$ were modeled as low pass IIR filters according to

$$H_1^{-1}(z) = 1 - 0.3z^{-1} - 0.1z^{-2}, \quad H_2^{-1}(z) = 1 - 0.2z^{-1}. \quad (29)$$

Both CS-LMS and NLMS algorithms use an eight tap weight vector initialized to zero and different step sizes (0.001, 0.01, and 0.1). The experimental setting described above models the conditions of the ANC in DXT systems except if only stationary signals are present. In order to show the performance of the filtering algorithms we use the MSE in the steady-state averaging over $J=200$ samples. The Monte Carlo simulation results (over 100 trials) of running the two algorithms are shown for different values of μ in Fig. 6. As shown in the figures the MSE of the CS-LMS algorithm is larger than the MSE of the NLMS over noise

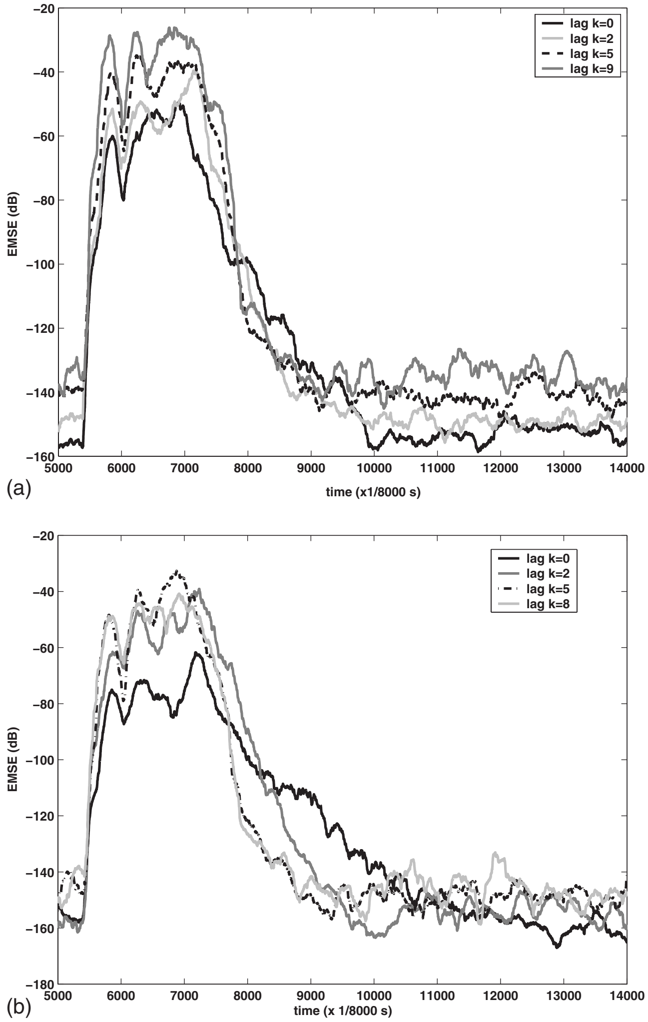

FIG. 5. ANC application: The trade-off between M and τ of the CS-LMS algorithm with increasing k over an utterance of the Spanish AURORA 3 database. (a) Evolution of EMSE [in dB=$20\log_e(.)$] for (μ=0.25, σ_v^2=0.3 L=12). (b) Evolution of EMSE [in dB=$20\log_e(.)$] for (μ=0.1, σ_v^2=0.5 L =12).

segments as expected [using Eq. (21) $\text{MSE}_{\text{LMS}} \simeq -29.8227$ dB; $\text{MSE}_{\text{CS-LMS}} \simeq -29.7170$ dB for $\mu=0.01$ and $\text{MSE}_{\text{LMS}} \simeq -28.4873$ dB; $\text{MSE}_{\text{CS-LMS}} \simeq -27.7641$ dB for $\mu=0.1$] [$\sigma_x^2=\sigma_v^2/(1-a^2(1)) < 4(\sigma_x^2-a(1))$, where $a(1)=0.2$ and $\sigma_x^2=1.0417$]. Note that $J_{\min}=-30$ dB on noise segments. Somehow when $s(n)$ turns on (to model correlated speech segments) there is a clear reduction in EMSE if μ is sufficiently high (to cope with high nonstationary environment). As also shown in theoretical section, on speech segments: $\text{MSE}_{\text{LMS}} \simeq 0.1932$ dB; $\text{MSE}_{\text{CS-LMS}} \simeq 0.0113$ dB for $\mu=0.01$ and $\text{MSE}_{\text{LMS}} \simeq 1.6012$ dB; $\text{MSE}_{\text{CS-LMS}} \simeq 0.074$ dB for $\mu=0.1$. Observe how when μ is increased the *small-step-size statistical theory* gives a rough result in the value of MSE comparing with the experimental one, although the relation between the performance of both algorithms is preserved. Note that the τ in CS-LMS is also larger than NLMS as expected.

B. ANC application: Stationary noise environment

Finally, to check the tracking ability of the set of algorithms we study the ANC problem using the previously defined AURORA 3 database. It is expected to obtain a better performance on the previous approaches since voice signals are nonstationary in nature. The simulations were carried out using recordings sampled at 8 kHz and processing 20 000 samples/recording (~2.5 s of real time). In the stationary case, the noise v was assumed to be zero mean white Gaussian with three different variances (σ_v^2=0.01, 0.1, 0.5) under severe noise condition as shown in Table I. The incoming clean signal was normalized using its entire span for convenience. In the simulations the filter length was varied in the set $L=\{8,12,24\}$, $\epsilon=0.0001$ and the constant step size was varied in the range $\mu=\{0.01,0.5\}$. The impulse response of the filters $\mathbf{h}_1$ and $\mathbf{h}_2$ were modeled as low pass IIR filters according to Eq. (29). Observe the effect of the equilibrium condition in the evaluation of the CS-LMS over an utterance of the Spanish AURORA 3 database in Fig. 7(a). The equilibrium condition can be easily achieved over the noise stationary segments; over the speech segments the equilibrium constraint is hardly satisfied because of the nonstationary nature of speech [regardless there is still an improvement in EMSE reduction—computed as $20\log_e(.)$ to highly decompress the EMSE—over classical LMS methods]. In the bottom of Fig. 7(a) we show the spectrogram of the estimated clean signal $e(n)$ by using the CS-LMS adaptation.

Table I shows the filtering results using the proposed and referenced algorithms over the complete Spanish AURORA 3 databases. As clearly demonstrated, the proposed method provides the minimum EMSE and M averaged over a set of filter lengths L for a wide range of noise variances (delimited by "\) except for the case σ_v^2=0.01, i.e., the EMSE of the NLMS is not affected by the step size μ since it is also proportional to the input power. In the case of nonstationary speech segments, this reduction of EMSE and M means better trade-off between convergence speed and misadjustment and then better filtering performance. The combination of the MLMS with the proposed CS-LMS using a VAD (as described in Sec. V) provides an improvement over the classical method MLMS operating separately at noise adverse conditions. Of course, the proposed CS-LMS still improves the combination of both using the VAD, since its more effective filtering on speech segments over MLMS. This behavior can be observed also in nonstationary experiments as shown in the following section. It is interesting to notice that for both environments, stationary and nonstationary, and for a wide range of parameters (L,μ,σ_v^2, etc.) we found a better performance over the referenced filtering algorithms.

Moreover, an additional improvement can be achieved by relaxing the equilibrium condition of the CS-LMS algorithm over noise segments and by employing the filtered energy-based VAD described in Sec. V B. The results using the SNR of 10 dB, which is the most probable application scenario for telecommunication terminals, are shown in Table II. The right combination of the lag-CS-LMS algorithm (k=5) and the standard CS-LMS outperforms all the references ANCs including the proposed CS-LMS working separately. Even if the combination of both ANCs is inversely used, that is the lag-LMS is applied to speech segments and the standard CS-LMS to noise periods, the results are still fair enough as shown in Table II.

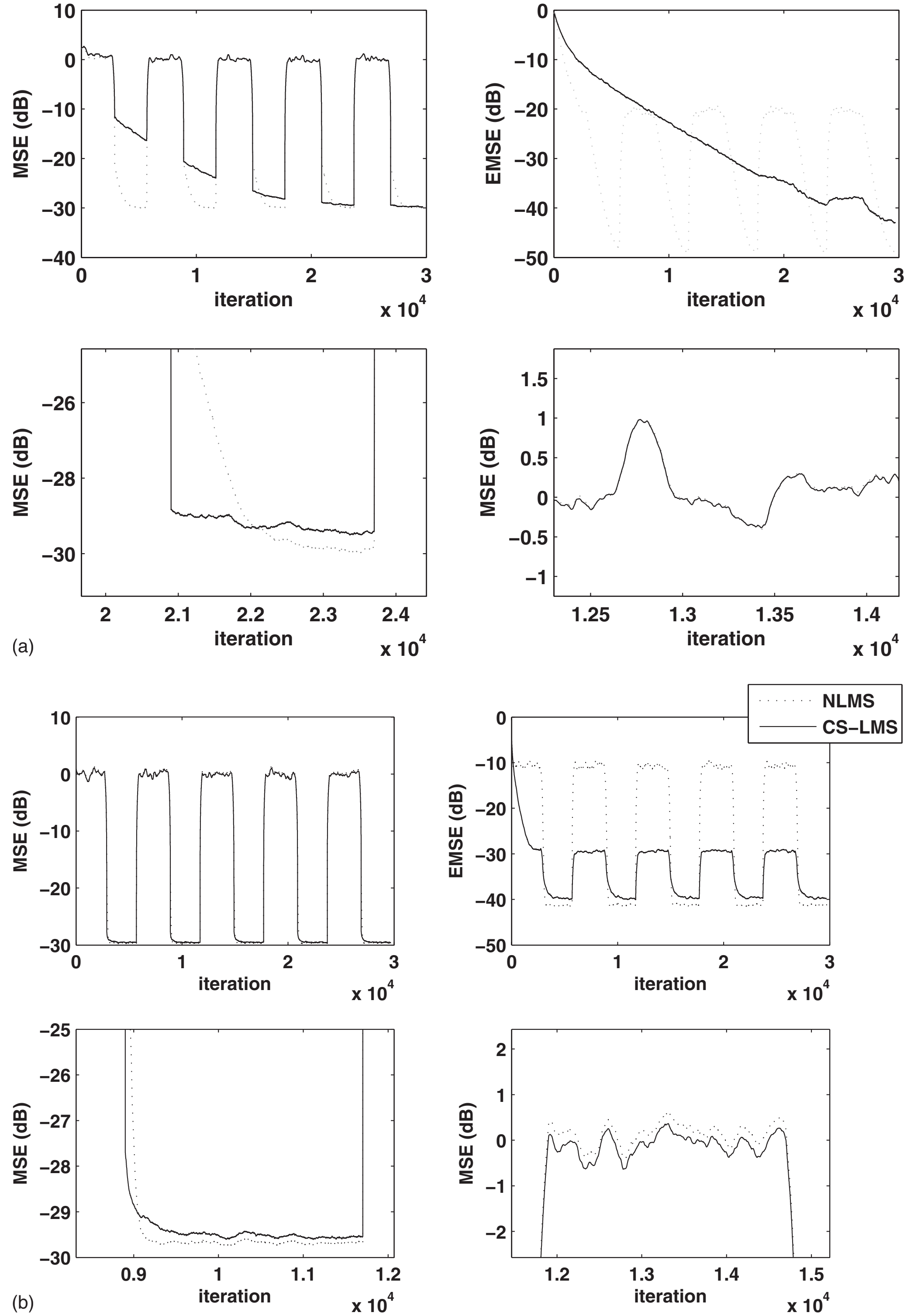

FIG. 6. Numerical experiment in the ANC problem. Top: MSE [in dB=10 $\log_{10}(.)$] and EMSE [in dB=20 $\log_{e}(.)$] comparison between the CS-LMS (line) and NLMS (dotted) algorithms. Bottom: zoom on MSE evolution over noise and speech segments. (a) Stationary environment (μ=0.01, σ_v^2=1 L=8). (b) Stationary environment (μ=0.1, σ_v^2=1 L=8).

C. ANC application: Nonstationary noise environment

For the nonstationary case the noise is also assumed to be zero mean Gaussian but with a variance linearly increasing from $\sigma^2_{v,\min}=0.1$ to $\sigma^2_{v,\max}=0.5$. The effect of increasing the noise variance is illustrated in Fig. 7(b). We also show the filtering result (spectrogram) of the CS-LMS algorithm. Table III shows the results obtained (EMSE and M) by the complete set of algorithms for different values of μ and L. Again, the proposed algorithm obtains the better trade-off between convergence speed and misadjustment although the computational burden is slightly increased by the novel adaptation. Obviously, the filtering performance is also affected by the nonstationarity of noise segments, unlike the stationary experiment described above, since a larger tracking abil-

TABLE I. Performance of referenced and proposed LMS algorithms in stationary environment.

Stationary white noise $E[v(n)]=0$ $\sigma_v^2=0.01/0.1/0.5$	NLMS		NDN-LMS		MLMS	
	$\overline{\text{EMSE}}$ (dB) $L=8;12;24$	$\bar{M}$ $L=8;12;24$	$\overline{\text{EMSE}}$ (dB) $L=8;12;24$	$\bar{M}$ $L=8;12;24$	$\overline{\text{EMSE}}$ (dB) $L=8;12;24$	$\bar{M}$ $L=8;12;24$
$\mu=0.1$	−42.23/−35.17/−23.42	0.22/0.43/1.61	−27.04/−26.15/−25.64	1.00/1.07/1.06	−47.26/−33.77/−22.88	0.11/0.56/1.82
$\mu=0.25$	−42.66/−31.80/−20.86	0.21/0.62/2.15	−22.75/−21.99/−21.94	1.57/1.67/1.67	−46.97/−30.31/−20.64	0.12/0.82/2.30
$\mu=0.5$	−42.91/−29.80/−19.31	0.20/0.85/2.56	−19.18/−18.85/−18.92	2.24/2.38/2.38	−45.47/−27.84/−19.25	0.14/1.07/2.65
Stationary white noise $E[v(n)]=0$ $\sigma_v^2=0.01/0.1/0.5$	EN-LMS		CS-LMS		VAD CS+MLMS	
	$\overline{\text{EMSE}}$ (dB) $L=8;12;24$	$\bar{M}$ $L=8;12;24$	$\overline{\text{ESME}}$ (dB) $L=8;12;24$	$\bar{M}$ $L=8;12;24$	$\overline{\text{EMSE}}$ (dB) $L=8;12;24$	$\bar{M}$ $L=8;12;24$
$\mu=0.1$	−41.90/−23.19/−6.95	0.23/1.85/13.25	−40.06/−39.36/−38.02	0.26/0.28/0.34	−45.29/−41.55/−38.02	0.11/0.22/0.34
$\mu=0.25$	−41.98/−21.02/−0.85	0.23/2.68/28.93	−36.50/−35.97/−35.71	0.39/0.41/0.43	−41.42/−38.09/−35.71	0.19/0.31/0.43
$\mu=0.5$	−41.70/−16.75/6.40	0.24/4.79/63.98	−33.49/−33.00/−32.91	0.56/0.59/0.59	−38.21/−35.21/−32.91	0.27/0.44/0.59

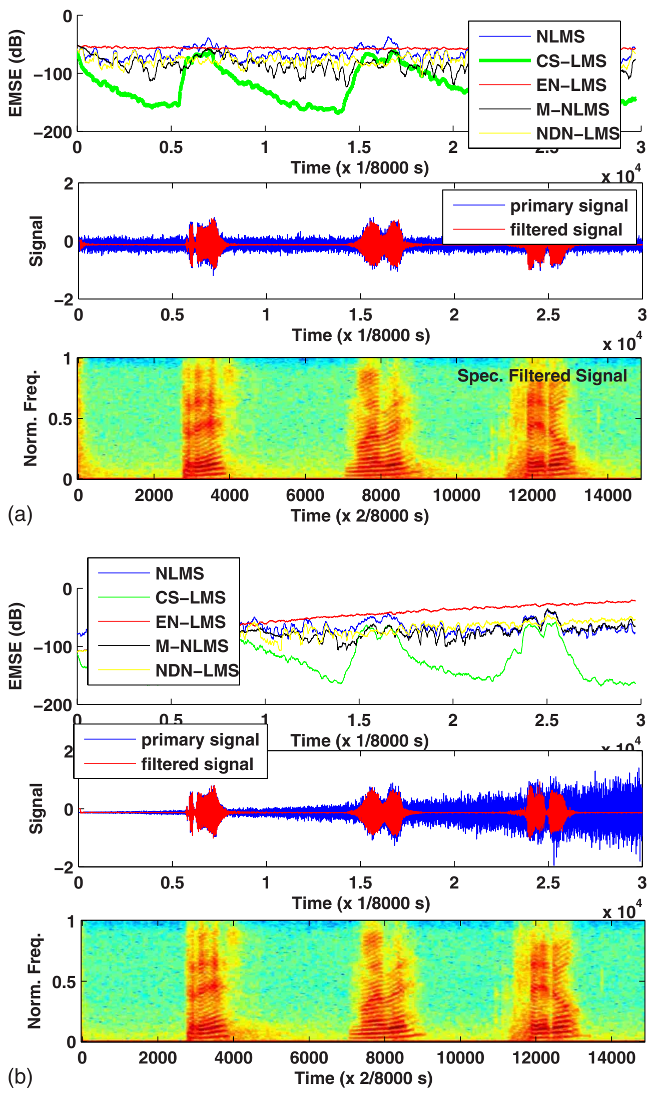

FIG. 7. (Color online) ANC application using the CS-LMS algorithm over an utterance of the Spanish AURORA 3 databases. Top: EMSE evolution [in dB$=20\log_e(.)$]; middle: corrupted and error (clean) signal; bottom: spectrogram of the error signal. (a) Stationary environment ($\mu=0.1$, $\sigma_v^2=0.1$ $L=12$). (b) Nonstationary environment ($\mu=0.1$, $\sigma_{v,\min}^2=0.1$, $\sigma_{v,\max}^2=0.7$, $L=12$).

ity is required during the whole utterance. The robustness of the proposed method under severe noise condition can be seen in Fig. 7(b). The above mentioned properties of our algorithm on the stationary scenario are still visible. Substantial improvements of the proposed ANC are also evident on this new scenario, particularly under conditions of strong noise. In addition, the combination of the selected ANC adaptations with a VAD again improves the filtering performance over the one obtainable with the referenced algorithms.

VII. CONCLUSION

In this paper we showed an ANC using the novel CS-LMS algorithm that is based on the minimization of the squared Euclidean norm of the change $\delta\mathbf{w}(n+1)$ subject to the constraint of equilibrium condition in the sequence of *a posteriori estimation errors*. To solve this constrained optimization problem, we used the well-known method of Lagrange multipliers for the general case of complex-valued data. We obtained a novel adaptation algorithm, which is a combination of a deterministic gradient function in order to minimize the MSE, with a gradient function maximizing the correlation function at lag 1 that was assumed to be real valued. Convergence analysis was studied in terms of the evolution natural modes toward the optimal Wiener–Hopf solution and convergence conditions were given: The stability performance depends *exclusively* on μ and the eigenvalues of $\Delta\mathbf{R}(1)$. In addition, we checked the benefits of a VAD scheme for improving the performance of the proposed algorithm. We combined the tracking ability and the higher performance on nonstationary speech segments of the latter algorithm for different lags under nonsevere noise conditions. For both stationary and nonstationary adverse noise environments, the proposed ANC based on the CS-LMS algorithm showed superior performance in decreasing excess MSE compared with referenced algorithms,[7,9,10] etc., using the AURORA 3 Spanish SDC database.[40]

ACKNOWLEDGMENTS

This work was partly supported by the Spanish Government under the PETRI DENCLASES (PET2006-0253),

TABLE II. Performance of referenced and proposed LMS algorithms in stationary environment.

Stationary white noise $E[v(n)]=0$ $\sigma_v^2=0.3$	NLMS		NDN-LMS		MLMS	
	EMSE (dB) $L=8$	M $L=8$	EMSE (dB) $L=8$	M $L=8$	EMSE (dB) $L=8$	M $L=8$
$\mu=0.01$	−36.60	0.38	−34.56	0.40	−34.29	0.55
Stationary white noise $E[v(n)]=0$ $\sigma_v^2=0.3$	Inv. VAD lag CS-LMS		CS-LMS		VAD lag CS-LMS	
	EMSE (dB) $L=8$	M $L=8$	EMSE (dB) $L=8$	M $L=8$	EMSE (dB) $L=8$	M $L=8$
$\mu=0.01$	−39.09	0.32	−40.18	0.28	−40.82	0.25

TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01) projects, and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, spain) under the Excellence Project (TIC-02566). The authors would like to thank Mrs. M. E. Cobo-Lara for help in writing this paper in English. J.M.G. would like to thank Mr. Monson Hayes for introducing him to the field of statistical signal processing.

APPENDIX A: PROOF OF THE CONSTRAINED-STABILITY LMS ADAPTATION

Once the Lagrangian function has been established in Eq. (6), we need to derive the minimum weight vector under the equilibrium constraint. Taking partial derivative of Eq. (6) with respect to the vector $\mathbf{w}^H(n+1)$ and set it equal to zero, we obtain

$$\frac{\partial\mathcal{L}(\mathbf{w}(n+1))}{\partial\mathbf{w}^H(n+1)} = \frac{\partial\delta\mathbf{w}^H(n+1)\delta\mathbf{w}(n+1)}{\partial\mathbf{w}^H(n+1)} + \lambda^*\left(\frac{\partial e^{[n+1]}(n)}{\partial\mathbf{w}^H(n+1)} - \frac{\partial e^{[n+1]}(n-1)}{\partial\mathbf{w}^H(n+1)}\right) = 0. \tag{A1}$$

Since $\delta\mathbf{w}(n+1)=\mathbf{w}(n+1)-\mathbf{w}(n)$ and $e^{[n+1]}(k)=d(k)-\mathbf{w}^H(n+1)\mathbf{x}(k)$ for $k=n,\ n-1$ then

$$\frac{\partial\mathcal{L}(\mathbf{w}(n+1))}{\partial\mathbf{w}^H(n+1)} = 2\,\delta\mathbf{w}(n+1) - \lambda^*\,\delta\mathbf{x}(n) = 0. \tag{A2}$$

thus, the step of the algorithm is

$$\delta\mathbf{w}(n+1) = \tfrac{1}{2}\lambda^*\,\delta\mathbf{x}(n) \Rightarrow \mathbf{w}(n+1) = \mathbf{w}(n) + \tfrac{1}{2}\lambda^*\,\delta\mathbf{x}(n). \tag{A3}$$

To complete the minimization procedure we must determine the Lagrange multiplier by applying the equilibrium constraint. Multiplying both sides of Eq. (A2) by $\delta\mathbf{x}^H(n+1)$ we have

$$2\,\delta\mathbf{x}^H(n)\,\delta\mathbf{w}(n+1) - \lambda^*\,\delta\mathbf{x}^H(n)\,\delta\mathbf{x}(n) = 0. \tag{A4}$$

Thus, we find that the Lagrange multiplier can be expressed as

$$\lambda^* = \frac{2\,\delta\mathbf{x}^H(n)\,\delta\mathbf{w}(n+1)}{\|\delta\mathbf{x}(n)\|^2} = \frac{2(\delta e^{[n+1]}(n) - \delta e^{[n]}(n))^*}{\|\delta\mathbf{x}(n)\|^2} \tag{A5}$$

since $e^{[k]}(n)=d(n)-\mathbf{x}^H(n)\mathbf{w}(k)$ and the numerator on the left side of Eq. (A5) is equal to $\mathbf{x}^H(n)\mathbf{w}(n+1)-\mathbf{x}^H(n-1)\mathbf{w}(n+1)-\mathbf{x}^H(n)\mathbf{w}(n)+\mathbf{x}^H(n-1)\mathbf{w}(n)$. Therefore applying the equilibrium constraint on the right hand side of Eq. (A5) ($\delta e^{[n+1]}(n)=0$), we have that λ is given by Eq. (7).

TABLE III. Performance of referenced and proposed LMS algorithms in nonstationary environment.

Nonstationary white noise $E[v(n)]=0$ $\sigma^2_{v,\text{min-max}}=0.1-0.5$	NLMS		NDN-LMS		MLMS	
	EMSE (dB) $L=8/12/24$	M $L=8/12/24$	EMSE (dB) $L=8/12/24$	M $L=8/12/24$	EMSE (dB) $L=8/12/24$	M $L=8/12/24$
$\mu=0.1$	−27.92/−26.27/−23.80	0.95/1.14/1.53	−24.72/−25.78/−27.01	1.09/1.07/1.05	−25.27/−25.16/−25.11	1.41/1.43/1.44
$\mu=0.25$	−24.70/−23.26/−21.13	1.37/1.62/2.07	−21.22/−22.01/−22.17	1.71/1.67/1.64	−22.65/−22.58/−22.49	1.86/1.88/1.89
$\mu=0.5$	−22.57/−21.32/−19.53	1.75/2.02/2.49	−18.70/−18.82/−19.07	2.43/2.38/2.34	−20.95/−20.91/−20.84	2.23/2.38/2.25
Nonstationary white noise $E[v(n)]=0$ $\sigma^2_{v,\text{min-max}}=0.1-0.5$	EN-LMS		CS-LMS		VAD CS+MLMS	
	EMSE (dB) $L=8/12/24$	M $L=8/12/24$	EMSE (dB) $L=8/12/24$	M $L=8/12/24$	EMSE (dB) $L=8/12/24$	M $L=8/12/24$
$\mu=0.1$	−16.49/−16.43/−16.44	3.30/3.35/3.35	−39.41/−38.79/−37.02	0.28/0.30/0.37	−26.97/−26.67/−26.98	1.26/1.29/1.33
$\mu=0.25$	−20.94/−20.80/−20.63	1.73/1.80/1.89	−35.97/−35.80/−35.14	0.42/0.42/0.42	−24.62/−24.32/−23.87	1.64/1.67/1.72
$\mu=0.5$	−23.61/−24.08/−24.18	1.31/1.28/1.23	−32.80/−32.92/−32.77	0.60/0.59/0.60	−23.10/−22.85/−22.41	1.92/1.96/2.02

APPENDIX B: PROOF OF THEOREM 1

One way to see this equivalence is to redefine the problem as follows. Let $\mathbf{x}_1(n)=\delta\mathbf{x}(n)$ be the difference incoming signal, $d_1(n)=\delta d(n)=d(n)-d(n-1)$ the difference desired signal, and $e_1(n)=\delta e(n)=d_1(n)-\mathbf{w}^H\mathbf{x}_1(n)$ the difference error signal (any other system variable will be denoted this way for short). Using this new set of transformed data and applying the Wiener–Hopf methodology, the classical filtering problem is to minimize the cost function

$$\mathbf{w}_{0,1}=\arg\min_{\mathbf{w}} E[|e_1(n)|^2]. \tag{B1}$$

Note that the gradient of the latter cost function defines the deterministic CS-LMS algorithm, that is, in terms of expected values. The optimum Wiener solution for the novel data set is then achieved at

$$\mathbf{w}_{0,1}=\mathbf{R}_{\mathbf{x}_1}^{-1}\mathbf{r}_{\mathbf{d}_1\mathbf{x}_1}, \tag{B2}$$

where the autocorrelation matrix $\mathbf{R}_{\mathbf{x}_1}=E[\mathbf{x}_1\mathbf{x}_1^H]$ is assumed to be positive definite and $\mathbf{r}_{\mathbf{d}_1\mathbf{x}_1}=E[\mathbf{x}_1 d_1^*]$ is the cross-correlation vector. The optimal solution $\mathbf{w}_{0,1}$ is indeed related, under some conditions, to the original optimal solution $\mathbf{w}_0$ given the generic data set. Generally, $\mathbf{w}_0\neq\mathbf{w}_{0,1}$, that is,

$$\mathbf{R}_{\mathbf{x}_1}^{-1}\mathbf{r}_{\mathbf{d}_1\mathbf{x}_1}\neq\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{r}_{\mathbf{dx}}. \tag{B3}$$

but if we assume that the desired signal is generated by the multiple linear regression model then

$$\mathbf{r}_{\mathbf{d}_1\mathbf{x}_1}=\mathbf{R}_{\mathbf{x}_1}\mathbf{w}_0\Rightarrow\mathbf{w}_0=\mathbf{w}_{0,1}. \tag{B4}$$

This can be also demonstrated in terms of the expected value of the difference error sequence in Eq. (B1) since

$$E[|e_1(n)|^2]=E[|(\mathbf{w}_0-\mathbf{w})^H\mathbf{x}_1(n)|^2]+E[|\delta e_0(n)|^2] \tag{B5}$$

is minimal only when $\mathbf{w}_{0,1}=\mathbf{w}_0$. In addition, once the optimal solution is achieved, the estimated desired signal $y_0(n)=\mathbf{w}_{0,1}^H\mathbf{x}(n)$ provides the minimum error sequence $e_{\min}=d(n)-y_0=e_0(n)$.

APPENDIX C: PROOF OF THE CONVERGENCE OF THE CS-LMS ALGORITHM

If the original tap-input vector $\mathbf{x}(n)$ and the desired response $d(n)$ are assumed to be jointly stationary, the correlation matrix and the cross-correlation vector of the difference input signal $\mathbf{x}_1(n)$ and the desired signal $d_1(n)$ can be expressed in terms of the original correlation matrix and cross-correlation vector

$$\mathbf{R}_{\mathbf{x}_1}=2\mathbf{R}-2\,\mathrm{Re}\{\mathbf{R}(1)\},$$

$$\mathbf{r}_{\mathbf{d}_1\mathbf{x}_1}=2\mathbf{r}_{\mathbf{dx}}(0)-2\,\mathrm{Re}\{\mathbf{r}_{\mathbf{dx}}(1)\}, \tag{C1}$$

where $\mathbf{R}(k)=E[\mathbf{x}(n+k)\mathbf{x}^H(n)]$ and $\mathbf{r}_{\mathbf{dx}}(k)=E[\mathbf{x}(n+k)d^*(n)]$, then the gradient vector in Eq. (14) is given by

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\hat{\mu}[2[\mathbf{r}_{\mathbf{dx}}(\mathbf{0})-\mathbf{R}(0)\mathbf{w}(n)] + 2\,\mathrm{Re}\{\mathbf{R}(1)\mathbf{w}(n)-\mathbf{r}_{\mathbf{dx}}(1)\}]. \tag{C2}$$

If real processes are also assumed, Eq. (C2) transforms into

$$\mathbf{w}(n+1)=\mathbf{w}(n)+2\hat{\mu}[\Delta\mathbf{r}_{\mathbf{dx}}(1)-\Delta\mathbf{R}(1)], \tag{C3}$$

where $\Delta\mathbf{r}_{\mathbf{dx}}(1)=\mathbf{r}_{\mathbf{dx}}(0)-\mathbf{r}_{\mathbf{dx}}(1)$ and $\Delta\mathbf{R}(1)=\mathbf{R}(0)-\mathbf{R}(1)$ is a Hermitian positive definite matrix. As in the SD algorithm the weight-error vector at time n can be defined as $\mathbf{c}(n)=\mathbf{w}_0-\mathbf{w}(n)$, where $\mathbf{w}_0$ is the optimal value of the tap-weight vector given by the Wiener–Hopf equations.[26] Then, eliminating the cross-correlation vector $\mathbf{r}_{\mathbf{dx}}(k)$ and rewriting the result in terms of the weight vector $\mathbf{c}(n)$, we get

$$\mathbf{c}(n+1)=(I-2\hat{\mu}\Delta\mathbf{R}(1))\mathbf{c}(n). \tag{C4}$$

In the latter expression we assume that $\mathbf{r}_{\mathbf{dx}}(1)=\mathbf{R}(1)\mathbf{w}_0$, that is, the desired signal is generated by a linear multiple regression model. Using eigendecomposition, we may express the difference correlation matrix as

$$\Delta\mathbf{R}(1)=\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H, \tag{C5}$$

where $\mathbf{Q}$ is the eigenvector matrix in columns and $\mathbf{\Lambda}$ is a diagonal matrix containing the real and non-negative eigenvalues $\{\lambda_1,\ldots,\lambda_n\}$. Thus, we can study the convergence analysis in terms of the evolution of each natural mode of the transformed form of Eq. (C4) as follows:

$$\mathbf{v}(n+1)=(I-2\hat{\mu}\mathbf{\Lambda})\mathbf{v}(n), \tag{C6}$$

where $\mathbf{v}(n+1)=\mathbf{Q}^H\mathbf{c}(n)$. For stability or convergence of the deterministic CS-LMS algorithm the following inequality must be satisfied:

$$-1<1-2\hat{\mu}\lambda_k<1 \quad \text{for all } k, \tag{C7}$$

where λ_k is the kth eigenvalue of the difference correlation matrix $\Delta\mathbf{R}(1)$. Since the eigenvalues are real and positive, it follows that a necessary and sufficient condition for convergence of the algorithm is that the step-size parameter $\hat{\mu}$ must satisfy the double inequality

$$0<\hat{\mu}<\frac{1}{\lambda_{\max}}, \tag{C8}$$

where $\lambda_{\max}$ is the largest eigenvalue of the difference correlation matrix $\Delta\mathbf{R}(1)$. If the convergence condition is compared with the one obtained using the SD algorithm, it can be seen that the latter is less restrictive by a factor of 2 than our algorithm (assuming that we have the same set of eigenvalues). It is also readily seen that an exponential envelope of the time constant τ_k can be fitted to the evolution of the natural modes, and its value, for the special case of slow adaptation (for small $\hat{\mu}$), is given by

$$\tau_k\simeq\frac{1}{2\hat{\mu}\lambda_k}, \quad \hat{\mu}\ll 1. \tag{C9}$$

In this case, the convergence time of each mode, using the same set of eigenvalues, is smaller than the natural modes of the SD algorithm by a factor of 2. As shown, the stability performance of the deterministic SC-LMS algorithm depends exclusively on μ and the eigenvalues of $\Delta\mathbf{R}(1)$. Note that a connection in terms of τ can be shown if we establish the following upper bound in the stability condition of the CS-LMS algorithm.

1. Stability condition for the stochastic CS-LMS algorithm

In the previous section we derived the stability condition for the deterministic CS-LMS algorithm, that is, the one that produces the adaptation in Eq. (8), when gradients are replaced with instantaneous estimates. It also follows from the NLMS analysis[26] that the high value of μ balances the trade-off between the M and the average time constant τ, which is larger than the one obtained with the standard LMS algorithm if $\mathrm{tr}(\mathbf{R}_{\mathbf{x}_1}) < \mathrm{tr}(\mathbf{R})$ since

$$\tau \simeq \frac{L}{\mu\, \mathrm{tr}(\mathbf{R}_{\mathbf{x}1})}, \tag{C10}$$

where L is the filter length. On the basis of this formula, we may make the same observations as in Ref. 26 about the connection between M and τ.

APPENDIX D: ON THE LEARNING CURVES OF THE STOCHASTIC CS-LMS ALGORITHM IN THE ANC APPLICATION

Using the definition of the weight-error vector $\varepsilon(n) = \mathbf{w}_0 - \mathbf{w}(n)$ and Eq. (8) with the step size defined as μ we may express the evolution of $\varepsilon(n)$ as

$$\varepsilon(n+1) = \varepsilon(n) - \mu \mathbf{x}_1(n)(s_1(n) + v_1(n) - (\mathbf{w}_0 - \varepsilon(n))^H \mathbf{X}_1(n))^*, \tag{D1}$$

where $v_1(n) = v(n) - v(n-1)$ and $v(n)$ denotes the noise in the primary signal $d(n)$. If we assume that $v(n)$ is generated by the multiple regression model: $v(n) = \mathbf{w}_0^H \mathbf{x}(n) + e_0(n)$ we find that the weight-error vector may be expressed as

$$\varepsilon(\nu+1) = (I - \mu \mathbf{x}_1(n)\mathbf{x}_1(n)^H)\varepsilon(n) - \mu \mathbf{x}_1(n)(\delta e_0(n) + s_1(n))^*, \tag{D2}$$

where $\delta \mathbf{x}(n)$ is denoted by $\mathbf{x}_1(n)$ for short. Invoking the direct-averaging method[26,37] we finally obtain Eq. (19). The stochastic evolution on the natural modes can be also studied transforming Eq. (19) into

$$\mathbf{v}(n+1) = (I - \mu \mathbf{\Lambda})\mathbf{v}(n) - \phi(n) \tag{D3}$$

by applying the unitary similarity transformation to the correlation matrix $\mathbf{R}_{\mathbf{x}_1}$, where $\mathbf{\Lambda} = \mathbf{Q}^H \mathbf{R}_{\mathbf{x}_1} \mathbf{Q}$ and the stochastic force vector is defined as $\phi(n) = \mu \mathbf{Q}^H \mathbf{x}_1 \delta\tilde{e}_0^*(n)$. It is seen readily that the latter vector has the following properties.

*P*1. The mean of the stochastic force vector $\phi(n)$ is zero: $E[\phi(n)] = \mathbf{0}$.

Proof: $E[\phi(n)] = \mathbf{Q}^H E[\mathbf{x}_1(n)\delta\tilde{e}_0^*(n)] = \mathbf{Q}^H E[\mathbf{x}_1(n)(\delta e_0(n) + s_1(n))^*] = 0$ by virtue of the principle of orthogonality[26] and statistical independence between input variables ($\mathbf{x}(n)$ and $s(n)$).

*P*2. The correlation matrix of the stochastic force vector is a diagonal matrix: $E[\phi(n)\phi^H(n)] = \mu^2 \tilde{J} \mathbf{\Lambda}$, where $\tilde{J} = 2(E[|e_0(n)^2|] + E[|s(n)^2|] - \mathrm{Re}\{E[s^*(n+1)s(n)]\})$.

Proof: Assuming $e_0(n)$ is a stationary and uncorrelated sequence and $s(n)$ is stationary,

$$E[\phi(n)\phi^H(n)] - \mu^2 \mathbf{Q}^H E[\mathbf{x}_1(n)\delta\tilde{e}_0^*(n)\delta\tilde{e}_0(n)\mathbf{x}_1^H(n)]\mathbf{Q} = \mu^2 \mathbf{Q}^H (E[\delta e_0^*(n)\delta e_0(n)] + E[s_1^*(n)s_1(n)])\mathbf{R}_{\mathbf{x}_1}\mathbf{Q}, \tag{D4}$$

which, by virtue of the principle of orthogonality and statistical independence, reduces to P2.

Using the properties P1 and P2 we obtain the same formulas for the first two moments of the natural modes $\mathbf{v}(n)$ as in Ref. 26, which allows to obtain the evolution of $J(n)$ with time step n. The second term of Eq. (18) in light of the direct-averaging method is equal to

$$\begin{aligned} &E[\varepsilon_0^H(n)\mathbf{x}(n)\mathbf{x}(n)^H \varepsilon_0(n)] \\ &\simeq E[\varepsilon_0^H(n)\mathbf{R}\varepsilon_0(n)] = \mathrm{tr}\{\mathbf{R}E[\varepsilon_0^H(n)\varepsilon_0(n)]\} \\ &= E\Big[\mathrm{tr}\Big\{\mathbf{v}^H \mathbf{Q}^H \Big(\tfrac{1}{2}\mathbf{R}_{\mathbf{x}_1} + \mathrm{Re}\{\mathbf{R}(1)\}\Big)\mathbf{Q}\mathbf{v}\Big\}\Big] \\ &= \tfrac{1}{2}\sum_{k=1}^{L} \lambda_k E[|v_k(n)|^2] + E[\mathrm{tr}\{\mathbf{v}^H \mathbf{Q}^H \mathrm{Re}\{\mathbf{R}(1)\}\mathbf{Q}\mathbf{v}\}]. \end{aligned} \tag{D5}$$

Assuming that the input signal is weakly correlated ($\mathbf{R}(1) \sim \mathbf{0}$) we may bound the second term in the last equality of Eq. (D5) with the first term (natural evolution), i.e., $E[\mathrm{tr}\{\mathbf{v}^H \mathbf{Q}^H \mathrm{Re}\{\mathbf{R}(1)\}\mathbf{Q}\mathbf{v}\}] \leqslant \frac{1}{2}\sum_{k=1}^{L}\lambda_k E[|v_k(n)|^2]$ then

$$J_{\mathrm{ex}}(n) \le \sum_{k=1}^{L} \lambda_k E[|v_k(n)|^2] = \sum_{k=1}^{L} \lambda_k \left(\frac{\mu\tilde{J}}{2 - \mu\lambda_k} + (1 - \mu\lambda_k)^{2n}\left(|v_k(0)|^2 - \frac{\mu\tilde{J}}{2 - \mu\lambda_k} \right) \right), \tag{D6}$$

where $v_k(n)$ denotes the component of natural mode $\mathbf{v}(n)$.[26] If the exponential factor is neglected with increasing n,

$$J_{\mathrm{ex}}(\infty) \le \sum_{k=1}^{L} \lambda_k \left(\frac{\mu\tilde{J}}{2 - \mu\lambda_k} \right) \simeq \frac{1}{2}\mu\tilde{J}\,\mathrm{tr}\{\mathbf{R}_{\mathbf{x}_1}\}. \tag{D7}$$

The reduction in $J_{\mathrm{ex}}(\infty)$ is achieved whenever

$$\begin{aligned} J_{\mathrm{ex}}(\infty) &= \tfrac{1}{2}\mu J\,\mathrm{tr}(\mathbf{R}_{\mathbf{x}_1}) \simeq \mu J\,\mathrm{tr}(\mathbf{R}) \le J_{\mathrm{ex}}^{\mathrm{LMS}}(\infty) \\ &= \tfrac{1}{2}\mu J_{\min}\,\mathrm{tr}(\mathbf{R}) \Leftrightarrow \mathrm{Re}\{r_s(1)\} \ge \tfrac{3}{4}J_{\min}, \end{aligned} \tag{D8}$$

where $r_s(1) = E[s^*(n+1)s(n)]$.

[1] J. Burg, "Application of the wiener filter theory to the design of seismic arrays," J. Acoust. Soc. Am. **33**, 1651–1684 (1961).

[2] D.-L. Liu and R. C. Waag, "Time-shift compensation of ultrasonic pulse focus degradation using least-mean-square error estimates of time delay," J. Acoust. Soc. Am. **92**, 2377–2378 (1992).

[3] D.-L. Liu and R. C. Waag, "Time-shift compensation of ultrasonic pulse focus degradation using least-mean-square error estimates of arrival time," J. Acoust. Soc. Am. **95**, 542–555 (1994).

[4] B. Widrow and S. D. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1985).

[5] B. Widrow, J. Glover, J. Mccool, J. Kaunitz, C. Williams, R. Hean, J. Zeidler, E. Dong, and R. Goodlin, "Adaptive noise cancelling: Principles and applications," Proc. IEEE **63**, 1696–1716 (1975).

[6] P. D. D. of neurodegenerative diseases using preprocessing techniques and classification algorithms, Reference No. PET2006-0253.

[7] J. Greenberg, "Modified lms algorithms for speech processing with an adaptive noise canceller," IEEE Trans. Speech Audio Process. **6**, 338–351 (1998).

[8]W. Harrison, J. S. Lim, and E. Singer, "A new application of adaptive noise cancellation," IEEE Trans. Acoust., Speech, Signal Process. **34**, 21–27 (1986).
[9]S. Douglas and T. H.-Y. Meng, "Stochastic gradient adaptation under general error criteria," IEEE Trans. Signal Process. **42**, 1335–1351 (1994).
[10]Z. Ramadan and A. Poularikas, "An adaptive noise canceller using error nonlinearities in the lms adaptation," in IEEE Proceedings of SoutheastCon (1979), Vol. **1**, pp. 359–364.
[11]M. Godavarti and I. A. O. Hero, "Partial update lms algorithms," IEEE Trans. Signal Process. **53**, 2382–2399 (2006).
[12]A. Osama and V. DeBrunner, "Hirschman optimal transform block lms adaptive filter," in IEEE International Conference on Acoustics, Speech and Signal Processing (2007), Vol. **3**, pp. 1305–1308.
[13]R. Bouquin-Jeannes and G. Faucon, "Study of a voice activity detector and its influence on a noise reduction system," Speech Commun. **16**, 245–254 (1995).
[14]J. M. Górriz, J. Ramírez, J. Segura, and C. Puntonet, "An effective cluster-based model for robust speech detection and speech recognition in noisy environments," J. Acoust. Soc. Am. **120**, 470–481 (2006).
[15]J. M. Górriz, J. Ramírez, J. Segura, and C. Puntonet, "Improved mo-lrt vad based on bispectra gaussian model," Electron. Lett. **41**, 877–879 (2005).
[16]ITU, "A silence compression scheme for G.729 optimized for terminals conforming to recommendation V.70," ITU-T Recommendation G.729-Annex B 1996.
[17]ETSI, "Voice activity detector (VAD) for adaptive multi-rate (AMR) speech traffic channels," ETSI EN 301 708 Recommendation, 1999.
[18]J. M. Górriz, J. Ramírez, E. Lang, and C. Puntonet, "Hard c-means clustering for voice activity detection," Speech Commun. **48**, 1638–1649 (2006).
[19]J. M. Górriz, J. Ramírez, C. Puntonet, and J. Segura, "Generalized lrt-based voice activity detector," IEEE Signal Process. Lett. **13**, 636–639 (2006).
[20]J. M. Górriz, J. Ramírez, and C. Puntonet, "An effective jointly pdf-based voice activity detector for real time applications," Electron. Lett. **3**, 251–253 (2007).
[21]J. Ramírez, J. M. Górriz, J. Segura, C. Puntonet, and A. Rubio, "Speech/non-speech discrimination based on contextual information integrated bispectrum lrt," IEEE Signal Process. Lett. **13**, 497–500 (2006).
[22]M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in Proceedings of the International Conference on Acoustics, Speech and Signal Processing (1979), Vol. **1**, pp. 208–211.
[23]S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120 (1979).
[24]S. Douglas, "A family of normalized lms algorithms," IEEE Signal Process. Lett. **1**, 49–51 (1994).
[25]N. J. Bershad, "Analysis of the normalized lms algorithm with gaussian inputs," IEEE Trans. Acoust., Speech, Signal Process. **34**, 793–806 (1986).
[26]S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1986).
[27]G. Goodwin and K. Sin, *Adaptive Filtering, Prediction and Control* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
[28]N. J. Bershad, "On the optimum data nonlinearity in lms adaptation," IEEE Trans. Acoust., Speech, Signal Process. **34**, 69–76 (1986).
[29]O. Tobias and R. Seara, "On the lms algorithm with constant and variable leakage factor in a nonlinear environment," IEEE Trans. Signal Process. **54**, 3448–3458 (2006).
[30]D. T. M. Slock, "On the convergence behavior of the lms and the normalized lms algorithms," IEEE Trans. Signal Process. **41**, 2811–2825 (1993).
[31]S. Choi, "Differential learning algorithms for decorrelation and independent component analysis," Neural Networks **19**, 1558–1567 (2006).
[32]D. Erdogmus, K. E. Hild, and J. C. Principe, "Online entropy manipulation: Stochastic information gradient," IEEE Signal Process. Lett. **10**, 242–245 (2003).
[33]D. Loftsgaarden and C. Queseberry, "A nonparametric estimate of a multivariate density function," Ann. Math. Stat. **36**, 1049–1051 (1965).
[34]G. Terrel and D. Scott, "Variable kernel density estimation," Ann. Math. Stat. **20**, 1236–1265 (1992).
[35]L. Breiman, W. Meisel, and E. Purcell, "Variable kernel estimates of probability densities," Technometrics **19**, 135–144 (1982).
[36]Y. Rao, D. Erdogmus, and J. Principe, "Error whitening criterion for adaptive filtering: Theory and algorithms," IEEE Trans. Signal Process. **53**, 1057–1069 (2005).
[37]H. Kushner, *Approximation and Weak Convergence Methods for Random Processes With Applications to Stochastic System Theory* (MIT, Cambridge, MA, 1984).
[38]J. Ramírez, J. M. Górriz, and J. Segura, "Statistical voice activity detection based on integrated bispectrum likelihood ratio tests for robust speech recognition," J. Acoust. Soc. Am. **121**, 2946–2958 (2007).
[39]M. Hayes, *Statistical Digital Signal Processing and Modeling* (Wiley, New York, 1996).
[40]A. Moreno, L. Borge, D. Christoph, R. Gael, C. Khalid, E. Stephan, and A. Jeffrey, "Speechdat-car: A large speech database for automotive environments," in Proceedings of the II LREC Conference (2000).