A Study of the ϵ-NLMS Algorithm With Application to Stereophonic Acoustic Echo Cancellation

Tetsuya Hoya, Jonathon A. Chambers and Patrick A. Naylor

Communications and Signal Processing Group,
Dept. of Electrical and Electronic Engineering,
Imperial College of Science, Technology and Medicine,
University of London, SW7 2BT, U.K.
e-mail: t.hoya@ic.ac.uk

Abstract

In Stereophonic Acoustic Echo Cancellation (SAEC), one of the fundamental problems lies in the misalignment in the filter coefficients due to the two strongly correlated channel-inputs. In this paper, we study the effect of the normalisation factor, ϵ, upon the convergence properties of the two-channel Normalised Least Mean Square (ϵ-NLMS) algorithm, through analysis and simulation studies with real speech datasets. We then show that the optimal choice for ϵ may be close to the variance of the channel-input data. Finally, a subband stereo echo canceller structure which uses a combination of the two-channel ϵ-NLMS and the Fast Least Squares (FLS) algorithms is proposed as a practical and promising solution to SAEC.

Key words: ϵ-NLMS algorithm, stereophonic acoustic echo cancellation, subband scheme

1 Introduction

In situations where spatial realism is desirable, as in teleconferencing, communications systems must have the potential to operate in a stereophonic mode. In such an environment, the use of stereophonic acoustic echo cancellers is necessary to reduce the undesirable echoes resulting from the coupling between the loudspeakers and the microphone [1]. Fig. 1 depicts the problem of Stereophonic Acoustic Echo Cancellation (SAEC). Due to the two strongly correlated channel-inputs, $x_1(n)$ and $x_2(n)$, adaptive algorithms suffer from misalignment in the filter coefficients, as well as slow convergence [1]. It is therefore important to understand how to adjust the parameters of such algorithms so as to minimise these effects.

Recently, two different NLMS type algorithms have been proposed for SAEC [2], namely the two-channel NLMS and the eXtended LMS (XLMS) algorithms. The latter algorithm can be viewed as a modified version of the two-channel NLMS algorithm for SAEC, which, to a first approximation, accounts for the inter-correlation between the two channel-inputs.

In practice, the NLMS type algorithms employ a small constant ϵ in the denominator within the normalisation term in order to avoid overflow in the division. In the single channel case, it has theoretically been shown that, unlike the original NLMS algorithm, the ϵ-NLMS algorithm has statistical behaviour dependent upon the input power level [3]. To date, several normalisation settings have been proposed around this scheme [4, 5, 6, 7]. In [5], the ϵ-NLMS algorithm is viewed in the form of a non-linearly modified LMS algorithm and both analytical and computer simulation studies on the basis of white input Gaussian data are provided. For the two-channel case, such a non-linearity may have some effect upon the convergence properties of an adaptive algorithm.

In this paper, we investigate the effect of the normalisation factor in full-band SAEC through both theoretical analysis and simulation study with real speech datasets (not the white input as assumed in analysis and used in some existing studies) and show that the optimal choice may be close to the variance of the channel-input data. In the simulation study of the subband SAEC [8, 9], we also show that, in comparison with the case where the two-channel Fast Least Squares (FLS) algorithm [2] is used in all frequency bands, a subband stereo echo canceller using a combination of the FLS algorithm in lower bands and the optimally-configured ϵ-NLMS algorithm in higher frequency bands gives significant improvement in terms of misalignment performance, while maintaining low computational complexity.

2 The Two-Channel ϵ-NLMS Type Algorithms

The update equation for the filter coefficients $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ at time index $k+1$ with the two-channel NLMS type algorithms is written in the form:

$$\begin{bmatrix} \boldsymbol{w}_1(k+1) \\ \boldsymbol{w}_2(k+1) \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1(k) \\ \boldsymbol{w}_2(k) \end{bmatrix} +$$

$$e(k) \left[\begin{array}{l} \boldsymbol{x}_1(k) f_1(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) \\ \boldsymbol{x}_2(k) f_2(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) \end{array} \right], \tag{1}$$

where $\boldsymbol{x}_i(k) = [x_{i,1}(k), x_{i,2}, ..., x_{i,(L-1)}(k)]^T$ $(i = 1, 2)$ are the respective tap input vectors of the twok channel-inputs and $e(k)$ is the error between the desired response $d(k)$ and sum value of the filter outputs:

$$e(k) \;=\; d(k) - \sum_{i=1}^{2} \boldsymbol{w}_i^T(k) \boldsymbol{x}_i(k), \tag{2}$$

and $(\cdot)^T$ denotes vector transpose.

2.1 Data Non-Linearity Within the ϵ-NLMS Algorithm

For the two-channel ϵ-NLMS algorithm $f_i(\cdot)$ $(i = 1, 2)$ are non-linear transformation functions given by

$$\begin{aligned} f_1(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) &= f_{\mathrm{NLMS},1}(k) \\ f_2(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) &= f_{\mathrm{NLMS},2}(k) \\ f_{\mathrm{NLMS},i}(k) &= \frac{\mu}{\epsilon_i + \|\boldsymbol{x}_1(k)\|_2^2 + \|\boldsymbol{x}_2(k)\|_2^2}, \end{aligned} \tag{3}$$

where μ is a learning constant, ϵ_i are small constants, and $\|\cdot\|_2^2$ denotes L_2 norm.

2.2 Analytical Evaluation of the ϵ_i Factor

For the analytical study of the ϵ_i factor within the two-channel ϵ-NLMS algorithm, we extend the work by S. C. Douglas and T. H. Y. Meng [5] to the two-channel case. In order that the convergence analysis of the two-channel ϵ-NLMS algorithm is mathematically tractable, the following assumptions are made:

Assumption 1:

The desired response $d(k)$ is given as:

$$d(k) = \sum_{i=1}^{2} \boldsymbol{h}_{i,opt}^T \boldsymbol{x}_i(k) + n(k) \tag{4}$$

where $\boldsymbol{h}_{i,opt}$ are the optimal coefficient vectors, and $n(k)$ is zero-mean uncorrelated noise signal [10] with variance σ_n^2 and is independent of the two-channel input signals $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$.

Assumption 2:

In order to utilise the independence assumption as in the single-channel case [11, 12], it is necessary to make the assumption that the two-channel input signals are uncorrelated, otherwise meaningful expressions can not be obtained. The data elements are either independent identically distributed (i.i.d.) with an even symmetric distribution or Gaussian with zero-means and known correlations [5].

Assumption 3:

The tap input vectors $\boldsymbol{x}_i(k)$ (with tap input vector lengths L) are independent of the corresponding previous data vectors $\boldsymbol{x}_i(k-p)$ $(p = 1, 2...)$.

Assumption 4:

The non-linear transformation functions $f_i(\cdot)$ are scalar functions of the sum of the L_2 norms of the tap input vectors $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$.

With the above four assumptions, we now move on to the mean and mean square behaviour of the two-channel ϵ-NLMS algorithm. We appreciate that Assumptions 2 and 3 are almost surely violated in practical SAEC, so the following results are only given as guidance.

2.2.1 Convergence in the Mean

We define the coefficient error matrix as

$$\begin{bmatrix} \boldsymbol{v}_1(k) \\ \boldsymbol{v}_2(k) \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1(k) \\ \boldsymbol{w}_2(k) \end{bmatrix} - \begin{bmatrix} \boldsymbol{h}_{1,opt} \\ \boldsymbol{h}_{2,opt} \end{bmatrix}, \tag{5}$$

using the expression above, we can re-write (2) as

$$e(k) = n(k) - \sum_{i=1}^{2} \boldsymbol{v}_i^T \boldsymbol{x}_i(k). \tag{6}$$

Now (1) is expressed as

$$\begin{aligned} \begin{bmatrix} \boldsymbol{v}_1(k+1) \\ \boldsymbol{v}_2(k+1) \end{bmatrix} &= \begin{bmatrix} \boldsymbol{v}_1(k) \\ \boldsymbol{v}_2(k) \end{bmatrix} + \\ &\{n(k) - \sum_{i=1}^{2} \boldsymbol{v}_i^T(k)\boldsymbol{x}_i(k)\} \begin{bmatrix} \boldsymbol{x}_1(k) f_1(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) \\ \boldsymbol{x}_2(k) f_2(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) \end{bmatrix}, \end{aligned} \tag{7}$$

taking the statistical expectation of both sides of (7) under the independence assumption, the expectations of the weight error and the data vectors are separated to give

$$\begin{aligned} \begin{bmatrix} E[\boldsymbol{v}_1(k+1)] \\ E[\boldsymbol{v}_2(k+1)] \end{bmatrix} &= \left\{ \boldsymbol{I}_{2L} - E \begin{bmatrix} \boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k) f_1(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) & \boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k) f_1(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) \\ \boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k) f_2(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) & \boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k) f_2(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) \end{bmatrix} \right\} \\ &\quad \begin{bmatrix} E[\boldsymbol{v}_1(k)] \\ E[\boldsymbol{v}_2(k)] \end{bmatrix}, \end{aligned} \tag{8}$$

where $\boldsymbol{I}_n$ is an n-by-n identity matrix.

Defining the unitary matrices $\boldsymbol{Q}_i$ $(i = 1, 2)$ and the diagonal matrices $\boldsymbol{\Lambda}_i$ of eigenvalues of covariances matrices $\boldsymbol{R}_i$, we may express the real-valued symmetric covariance matrices as

$$\boldsymbol{R}_i = \boldsymbol{Q}_i \boldsymbol{\Lambda}_i \boldsymbol{Q}_i^T$$

Then, we express (8) in the transformed domain using the definitions of the rotated vectors $\tilde{\boldsymbol{v}}_i = \boldsymbol{Q}_i^T \boldsymbol{v}_i$ and $\tilde{\boldsymbol{x}}_i = \boldsymbol{Q}_i^T \boldsymbol{x}_i$:

$$\begin{bmatrix} E[\tilde{\boldsymbol{v}}_1(k+1)] \\ E[\tilde{\boldsymbol{v}}_2(k+1)] \end{bmatrix} = \left\{ \boldsymbol{I}_{2L} - E \begin{bmatrix} \tilde{\boldsymbol{x}}_1(k)\tilde{\boldsymbol{x}}_1^T(k) f_1(\tilde{\boldsymbol{x}}_1(k), \tilde{\boldsymbol{x}}_2(k)) & \tilde{\boldsymbol{x}}_1(k)\tilde{\boldsymbol{x}}_2^T(k) f_1(\tilde{\boldsymbol{x}}_1(k), \tilde{\boldsymbol{x}}_2(k)) \\ \tilde{\boldsymbol{x}}_2(k)\tilde{\boldsymbol{x}}_1^T(k) f_2(\tilde{\boldsymbol{x}}_1(k), \tilde{\boldsymbol{x}}_2(k)) & \tilde{\boldsymbol{x}}_2(k)\tilde{\boldsymbol{x}}_2^T(k) f_2(\tilde{\boldsymbol{x}}_1(k), \tilde{\boldsymbol{x}}_2(k)) \end{bmatrix} \right\} \begin{bmatrix} E[\tilde{\boldsymbol{v}}_1(k)] \\ E[\tilde{\boldsymbol{v}}_2(k)] \end{bmatrix}, \tag{9}$$

where $f_i(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2) = f_i(\boldsymbol{x}_1, \boldsymbol{x}_2)$, because of the special forms of $f_i(\cdot)$.

If we set $f(\cdot) = f_1(\cdot) = f_2(\cdot)$ with the assumptions in [5] and assume that the two-channel input signals are uncorrelated, we obtain the stability condition in the mean sense given in the form:

$$0 < \mu < \frac{2}{\max_{i=1,2, j=1,2,\ldots,L} E[\tilde{x}_{i,j}^2(k) f(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k))]}, \tag{10}$$

in which $\tilde{x}_{i,j}$ is the j-th element of the rotated vector $\tilde{\boldsymbol{x}}_i$. Moreover, the time constants of adaptation for the means of the weight vector elements are given by

$$\tau_{M,i,j} = \frac{-1}{ln(1 - \mu E[\tilde{x}_{i,j}^2(k) f(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k))])}. \tag{11}$$

If we use a combined form for the two rotated tap input vectors as $\tilde{\boldsymbol{x}}(k) = [\tilde{\boldsymbol{x}}_1^T(k)\ \tilde{\boldsymbol{x}}_2^T(k)]^T$, the non-linear transformation function $f(\cdot)$ becomes identical to that in [5], and hence both the expressions (10) and (11) reduce to those obtained for the single-channel case in [5].

2.2.2 Convergence in the Mean Square

To determine the mean square behaviour of the two-channel ϵ-NLMS algorithm, we post-multiply both sides of (7) by its transpose, dropping the arguments in $f_i(\cdot)$ for notational convenience, leading to

$$\begin{aligned} \begin{bmatrix} \boldsymbol{v}_1(k+1) \\ \boldsymbol{v}_2(k+1) \end{bmatrix} \cdot & \left[\boldsymbol{v}_1^T(k+1)\ \boldsymbol{v}_2^T(k+1)\right] = \\ & \left\{ \boldsymbol{I}_{2L} - \mu \begin{bmatrix} \boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k) f_1 & \boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k) f_1 \\ \boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k) f_2 & \boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k) f_2 \end{bmatrix} \right\} \begin{bmatrix} \boldsymbol{v}_1(k) \\ \boldsymbol{v}_2(k) \end{bmatrix} \cdot \left[\boldsymbol{v}_1^T(k)\ \boldsymbol{v}_2^T(k)\right] \\ & \cdot \left\{ \boldsymbol{I}_{2L} - \mu \begin{bmatrix} \boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k) f_1 & \boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k) f_2 \\ \boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k) f_1 & \boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k) f_2 \end{bmatrix} \right\} \\ & + \mu^2 n^2(k) \begin{bmatrix} \boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k) f_1^2 & \boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k) f_1 f_2 \\ \boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k) f_1 f_2 & \boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k) f_2^2 \end{bmatrix} \\ & + \mu n(k) \begin{bmatrix} \boldsymbol{x}_1(k) f_1 \\ \boldsymbol{x}_2(k) f_2 \end{bmatrix} \left[\boldsymbol{v}_1^T(k)\ \boldsymbol{v}_2^T(k)\right] \\ & \cdot \left\{ \boldsymbol{I}_{2L} - \mu \begin{bmatrix} \boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k) f_1 & \boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k) f_2 \\ \boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k) f_1 & \boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k) f_2 \end{bmatrix} \right\} \end{aligned}$$

$$+\mu n(k)\left\{\boldsymbol{I}_{2L}-\mu\begin{bmatrix}\boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k)f_1 & \boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k)f_1\\ \boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k)f_2 & \boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k)f_2\end{bmatrix}\right\}$$
$$\cdot\begin{bmatrix}\boldsymbol{v}_1(k)\\ \boldsymbol{v}_2(k)\end{bmatrix}\left[\boldsymbol{x}_1^T(k)f_1\ \boldsymbol{x}_2^T(k)f_2\right]. \tag{12}$$

Taking the statistical expectation of both sides of (12) under Assumption 1, we have

$$\begin{bmatrix}E[\boldsymbol{v}_1(k+1)\boldsymbol{v}_1^T(k+1)] & E[\boldsymbol{v}_1(k+1)\boldsymbol{v}_2^T(k+1)]\\ E[\boldsymbol{v}_2(k+1)\boldsymbol{v}_1^T(k+1)] & E[\boldsymbol{v}_2(k+1)\boldsymbol{v}_2^T(k+1)]\end{bmatrix}=\begin{bmatrix}E[\boldsymbol{v}_1(k)\boldsymbol{v}_1^T(k)] & E[\boldsymbol{v}_1(k)\boldsymbol{v}_2^T(k)]\\ E[\boldsymbol{v}_2(k)\boldsymbol{v}_1^T(k)] & E[\boldsymbol{v}_2(k)\boldsymbol{v}_2^T(k)]\end{bmatrix}$$
$$-\mu\left\{\begin{bmatrix}E[\boldsymbol{v}_1(k)\boldsymbol{v}_1^T(k)] & E[\boldsymbol{v}_1(k)\boldsymbol{v}_2^T(k)]\\ E[\boldsymbol{v}_2(k)\boldsymbol{v}_1^T(k)] & E[\boldsymbol{v}_2(k)\boldsymbol{v}_2^T(k)]\end{bmatrix}\cdot\begin{bmatrix}E[\boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k)f_1] & E[\boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k)f_1]\\ E[\boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k)f_2] & E[\boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k)f_2]\end{bmatrix}\right.$$
$$\left.+\begin{bmatrix}E[\boldsymbol{v}_1(k)\boldsymbol{v}_1^T(k)] & E[\boldsymbol{v}_2(k)\boldsymbol{v}_1^T(k)]\\ E[\boldsymbol{v}_1(k)\boldsymbol{v}_2^T(k)] & E[\boldsymbol{v}_2(k)\boldsymbol{v}_2^T(k)]\end{bmatrix}\cdot\begin{bmatrix}E[\boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k)f_1] & E[\boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k)f_2]\\ E[\boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k)f_1] & E[\boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k)f_2]\end{bmatrix}\right\}$$
$$+\mu^2\begin{bmatrix}E[\boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k)f_1] & E[\boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k)f_1]\\ E[\boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k)f_2] & E[\boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k)f_2]\end{bmatrix}\begin{bmatrix}E[\boldsymbol{v}_1(k)\boldsymbol{v}_1^T(k)] & E[\boldsymbol{v}_1(k)\boldsymbol{v}_2^T(k)]\\ E[\boldsymbol{v}_2(k)\boldsymbol{v}_1^T(k)] & E[\boldsymbol{v}_2(k)\boldsymbol{v}_2^T(k)]\end{bmatrix}$$
$$\cdot\begin{bmatrix}E[\boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k)f_1] & E[\boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k)f_2]\\ E[\boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k)f_1] & E[\boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k)f_2]\end{bmatrix}$$
$$+\mu^2\sigma_n^2\begin{bmatrix}E[\boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k)f_1^2] & E[\boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k)f_1f_2]\\ E[\boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k)f_1f_2] & E[\boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k)f_2^2]\end{bmatrix}. \tag{13}$$

Note that we cannot proceed without Assumption 2.

Following the way we analysed the convergence in the mean sense, using a combined form for the tap input vectors $\boldsymbol{x}(k)=[\boldsymbol{x}_1^T(k)\ \boldsymbol{x}_2^T(k)]^T$ in order to fix the non-linear transformation functions as $f_i(\cdot)=f(\cdot)$, we obtain an essentially identical form for the misadjustment in [5] under the independence assumption, that is:

$$M=\frac{2\mu L\lambda E[\tilde{x}^2f^2(\tilde{\boldsymbol{x}})]}{2E[\tilde{x}^2f^2(\tilde{\boldsymbol{x}})]-\mu E[\tilde{x}^2\|\tilde{\boldsymbol{x}}\|_2^2f^2(\tilde{\boldsymbol{x}})]}, \tag{14}$$

in which the elements of the combined tap input vector in $\tilde{\boldsymbol{x}}$, $\tilde{x}_j$ $(j=1,2,...,2L)$ can be set to a single value $\tilde{x}$ and so can the elements in the data covariance matrix, i.e., $\lambda_j{=}\lambda$ (note that the iteration index k is dropped for $\tilde{\boldsymbol{x}}$). As expected, the misadjustment is dependent upon μ, L, λ, and $f(\tilde{\boldsymbol{x}})$. The choice of $f(\cdot)$ may therefore have impact upon the convergence properties of the two-channel NLMS algorithm.

2.2.3 Derivation of the Non-Linear Transformation Functions for the Two-Channel Case

Using the combined form for both the two tap input and the weight error vectors, as in the mean and the mean square analyses, the derivation of the non-linear transformation functions for the two-channel case also follow the same approach as in [5]. The non-linear transformation function derived in [5] is given by

$$f(k)=\frac{1}{1/\mu+\|\boldsymbol{x}(k)\|_2^2}. \tag{15}$$

For the two-channel case, we can re-write the above expression as a modification to (3):

$$f'_{\mathrm{NLMS},i}(k) = \frac{1}{1/\mu + \|\boldsymbol{x}_1(k)\|_2 + \|\boldsymbol{x}_2(k)\|_2^2} \ (i = 1, 2). \tag{16}$$

The effective range of $1/\mu$ in (16) is, however, still not known [5], and to obtain a specific convergence rate or final mean-square error is a non-trivial task [3]. However, as we will show in the experimental evaluation described next, the misalignment performance can be greatly improved with appropriate settings of ϵ_i in (3) for the two-channel case.

3 Simulation Study

In the simulation study, we firstly study the effect of the normalisation factor in full-band SAEC and its optimal setting is evaluated by experimentation. The optimal setting is then exploited and applied to subband SAEC.

3.1 Parameter Settings

In the simulation, to represent the receiving room impulse responses, models based upon zero-mean Gaussian random variables modulated by exponentially decaying/growing envelopes were used [14] and an example of which is shown in Fig. 2. The synthetic filter coefficients were generated and modulated with variations simulated by a "random-walk" regression model.

The input data used were the real speech data corresponding to the following five sentences:

"Pleasant zoos are rarely reached by efficient transportation."
"Cherry muffins are nice, but blueberry ones are nicer."
"The rain in Spain falls mainly on the plains."
"The quick brown fox jumps over the lazy dog."
"How are you, you early owl?"

Each sentence was recorded by five (four, for the last sentence) different speakers (i.e., two female and three male speakers, and in total 24 different utterances were used) in a quiet room, sampled originally at 48kHz, and down-sampled to 8kHz. The input data were then normalised to have a maximum magnitude unity.

In order to evaluate the performance, two measurements, i.e., the segmental Echo Return Loss Enhancement (ERLE) and the misalignment, are considered in this paper.

The segmental Echo Loss Return Enhancement (ERLE) at frame j is given by

$$\mathrm{ERLE}(j) \ = \ 10\log_{10} \sum_{i=0}^{255} \frac{d^2(256j+i)}{e^2(256j+i)}. \tag{17}$$

In full-band SAEC, the performance is evaluated based upon the misalignment between the filter coefficient vectors $\boldsymbol{w}_i(k)$ $(i = 1, 2)$ and the optimal $\boldsymbol{h}_i$:

$$\text{Misalignment} = 10\log_{10}\frac{1}{2}\sum_{i=1}^{2}\frac{\|\boldsymbol{w}_i(k) - \boldsymbol{h}_i\|_2}{\|\boldsymbol{h}_i\|_2}. \tag{18}$$

In contrast, in subband SAEC, the measurement based upon the averaged misalignment is used [8]. The averaged misalignment between the filter coefficient vectors in the jth frequency band $\boldsymbol{w}_i^j(k)$ $(i = 1, 2,\ j = 1, 2, ..., N)$ and the optimum $\boldsymbol{h}_i^j$ is computed as

$$\text{Averaged Misalignment}(k) = 10\log_{10}\frac{1}{2N}\sum_{i=1}^{2}\sum_{j=1}^{N}\frac{\|\boldsymbol{w}_i^j(k) - \boldsymbol{h}_i^j\|_2}{\|\boldsymbol{h}_i^j\|_2}. \tag{19}$$

3.2 Experimental Evaluation for the Optimal Normalisation Setting in Full-Band SAEC

In the evaluation study for the optimal normalisation setting, the ϵ_i $(i = 1, 2)$ values were kept identical for each channel, i.e., $\epsilon_i = \epsilon$. Firstly, ϵ was set to 1, 0.1, 0.01, and 0 (with due consideration of the channel-input power). For the setting $\epsilon = 0$ with the power consideration, no weight update was performed when the power of the tap-input vector is below a given threshold. The threshold value was determined by calculating the energy of the first L samples, where L is the tap input vector length. The modification to (3) given by (16) was then used. For all the above settings, the learning constant μ was fixed to 0.9.

Since the obtained performance order is nearly the same, we only present the results using two different speech signals in this paper. Figs. 3 and 4 respectively show the misalignment performance obtained for the two different speech signals. In these figures, various values of ϵ are compared (including the setting given by (16)) in the ϵ-NLMS algorithm using $L = 50$. The results are shown in the presence of noise in the echo-path in the receiving room at SNR=30dB. The noise in the echo-path was assumed as an independent Gaussian random noise signal. It is clear that over 20,000 samples there is advantage in particular setting of ϵ. The performance order for different settings of ϵ remained constant for longer simulation intervals due, we expect, to the time variation in the optimal echo path coefficients.

3.3 Discussion on the Evaluation Study

As shown in Figs. 3 and 4, the difference in terms of the misalignment performance between the normalisation settings is substantial at later iterations. In the evaluation study, we also observed that the performance difference becomes smaller as the tap input vector length increases.

These simulations suggest that the performance depends upon (i) channel input power and (ii) filter length. The dependence upon input power is consistent with the findings of [3] in which a loose bound on the optimal range of ϵ is given as $\epsilon << $ variance of the input

signal. From tests performed in this study, however, it has been observed that the best performance was obtained when ϵ was set *close to the variance of the signal.* In particular, the finding has been confirmed in a separate experiment for a speech signal scaled to have unit variance using ϵ-NLMS with L=50.

The dependence upon filter length comes about since, as L increases, the estimate values of $\|\boldsymbol{x}_1\|_2$ and $\|\boldsymbol{x}_2\|_2$ govern the adaptation gain with ϵ becoming insignificant.

In further studies, we also observed that the variations between the ϵ-NLMS algorithm with different normalisation factor settings and the FLS algorithm [2] becomes more apparent as the filter length increases. This indicates that the effective condition number of the input correlation matrix in SAEC becomes much more significant in terms of the performance as the number of the taps is increased, as predicted by the bordering theorem of symmetric matrices [16]

In practice, however, there must be trade-off; although the overall performance would be greatly improved at long filter lengths by employing least squares type algorithms such as FLS, the occurrence of a numerical instability as well as the increase in computational complexity will be problematic, especially during intervals when the variance of the channel-input is very small (in fact, it is known that the performance of the FLS algorithm [2] is very dependent upon the forgetting factor γ, as demonstrated in Fig. 5).

This fact motivates the consideration of subband approaches with the optimally-configured NLMS type algorithms within SAEC, as described next.

3.4 Experiments Using A Subband Stereo Echo Canceller

The structure of the subband stereo echo canceller used in this paper is depicted in Fig. 6. In the figure, $\boldsymbol{w}^j$ denotes the estimated filter coefficients vector in the ith subband and $\boldsymbol{h}^j$ the corresponding vector of the modelled receiving room impulse response. We emphasise that this subband model of the true echo path is only an approximation.

In the experiment, the lengths of the modelled receiving room impulse responses $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in full-band were assumed to be $L = 1024$, and the signals are divided into $N = 8$ frequency subbands using a three-stage binary tree structure of halfband systems (i.e., the filter length in subband i is accordingly set to $L_i = 128$). The filter banks employed are 32-tap FIR QMFs [15]. As for the full-band case, the noise in the receiving room was set to SNR=30dB.

Fig. 7 shows a comparison of the segmental ERLE performance. The misalignment performance obtained is shown in Fig. 8.

A performance comparison is made between the subband stereophonic acoustic echo canceller structure using the FLS algorithm in all the subbands and using the FLS algorithm in the first two lowest frequency bands (i.e., below 1kHz) and the optimally configured ϵ-NLMS algorithm in the remaining higher bands.

For the ϵ-NLMS algorithm, ϵ_i in (3) were identically set to the averaged variance of the two channel-inputs over the 24 different speech samples.

For the FLS algorithm, the performance can be improved further by setting the forgetting factor individually in each subband. However, such selection must be speech-

dependent and often causes numerical instability problem (in fact, this was confirmed during the simulation: 8 utterances out of 24 were rejected due to the numerical instability). For these simulations, the forgetting factor of the FLS algorithm was *identically* set to 0.999 in each subband. The value was determined so that no rejection occurred over the 24 utterances.

4 Conclusion

In this paper, the utility of the two-channel ϵ-NLMS algorithm for SAEC has been studied both analytically and on the basis of simulations with real speech signals. From the experimental evaluation, it has been shown that the ϵ-NLMS algorithm with the optimal normalisation settings can significantly improve the misalignment performance and that ϵ_i should be selected close to the variance of the channel-inputs. In the analytical study, the non-linear functions for the two-channel case have been derived under rather strong assumptions due to the mathematical difficulty. It remains as an open question as to whether the performance improvement observed is simply due to normalisation or a consequence of the non-linear transformation of the input data contained in (3). In the simulation study, the results obtained by a subband stereo echo canceller using the optimally configured ϵ-NLMS algorithm in the higher frequency bands show around 4-5dB misalignment performance improvement, while requiring less computational complexity ($O(6 \times 4L + 2 \times 28L)$) than the echo canceller using the FLS algorithm in all the subbands ($O(8 \times 28L)$).

References

[1] M. M. Sondhi, D. R. Morgan, and J. Hall, "Stereophonic Acoustic Echo Cancellation", IEEE Signal Processing Letters, Vol. 2, No. 8, pp. 148-151, Aug. 1995.

[2] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier, "Adaptive Filtering Algorithms For Stereophonic Acoustic Echo Cancellation", Proc. ICASSP-96, Vol. 5, pp. 3099-3102, Atlanta, May, 1996.

[3] N. J. Bershad, "Behavior of the ϵ-Normalized LMS Algorithm with Gaussian Inputs", IEEE Trans. ASSP, Vol. ASSP-35, No. 5, pp. 636-644, May 1987.

[4] N. J. Bershad, "On the Optimum Data Nonlinearity in LMS Adaptation", IEEE Trans. ASSP, Vol. ASSP-34, No. 1, pp. 69-76, Feb. 1986.

[5] S. C. Douglas and T. H. Y. Meng, "Normalized Data Nonlinearities for LMS Adaptation", IEEE Trans. SP, Vol. 42, No. 6, pp. 1352-1365, June 1994.

[6] J. E. Greenberg, "Modified LMS Algorithms for Speech Processing with an Adaptive Noise Canceller", IEEE Trans. on SAP, Vol. 6, No. 4, pp. 338-351, July 1998.

[7] D. T. M. Slock, "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms", IEEE Trans. SP, Vol. 41, No. 9, pp. 2811-2825, Sept. 1993.

[8] J. Benesty, D. R. Morgan, J. L. Hall, and M. M. Sondhi, "Stereophonic Acoustic Echo Cancellation Using Nonlinear Transformations And Comb Filtering", Proc. ICASSP-98, Vol. VI, pp. 3673-3676, Seattle, May 1998.

[9] S. Makino, K. Strauss, S. Shimauchi, Y. Haneda, and A. Nakagawa, "Subband Stereo Echo Canceller Using the Projection Algorithm With Fast Convergence to the True Echo Path", Proc. ICASSP-97, Vol. I, pp. 299-302, Munich, Apr. 1997.

[10] S. C. Douglas, "Performance Comparison of Two Implementations of the Leaky LMS Adaptive Filter", IEEE Trans. on SP, Vol. 45, No. 8, pp. 2125-2129, Aug. 1997.

[11] W. A. Gardner, "Learning Characteristics of Stochastic-Gradient-Descent Algorithms: A General Study, Analysis, and Critique", Signal Processing, Vol. 6, No. 2, pp. 119-133, Apr. 1984.

[12] J. E. Mazo, "On the Independence Theory of Equalizer Convergence", Bell Syst. Tech. J., Vol. 58, No. 3, pp. 963-993, May-June, 1979.

[13] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation", Proc. ICASSP'97, Vol. 1, pp.303-306, Munich, Apr., 1997.

[14] T. Hoya, J. A. Chambers, N. Forsyth, and P. A. Naylor, "Steady-State Solutions of the Extended LMS Algorithm for Stereophonic Acoustic Echo Cancellation With Leakage or Signal Conditioning", Proc. EUSIPCO-98, Vol. 2, pp. 977-980, Rhodes, Sept. 1998.

[15] R. E. Crochiere and L. R. Rabiner, "Multirate Digital Signal Processing", Englewood Cliffs, NJ: Prentice-Hall, 1983.

[16] M. H. Hayes, "Statistical Digital Signal Processing and Modeling", John Wiley & Son, Inc., 1996.

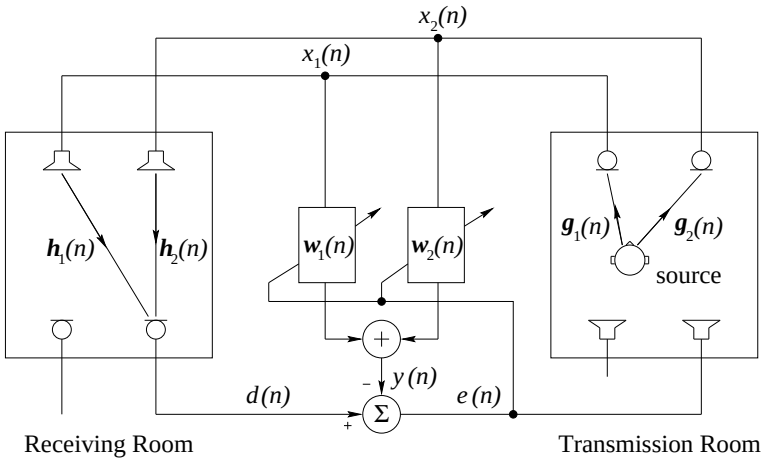

Figure 1: Schematic Diagram of Stereophonic Acoustic Echo Cancellation (only a single-channel echo-canceller is depicted for clarity)

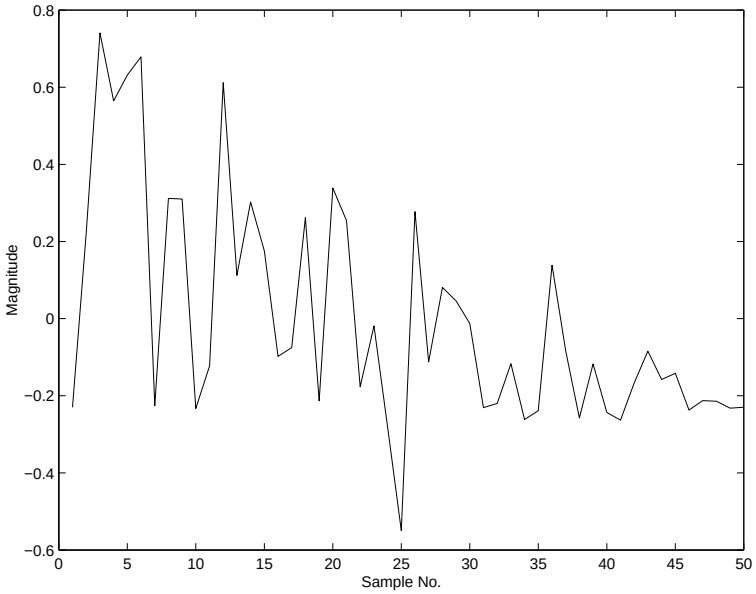

Figure 2: An Example of Modelled Room Impulse Response

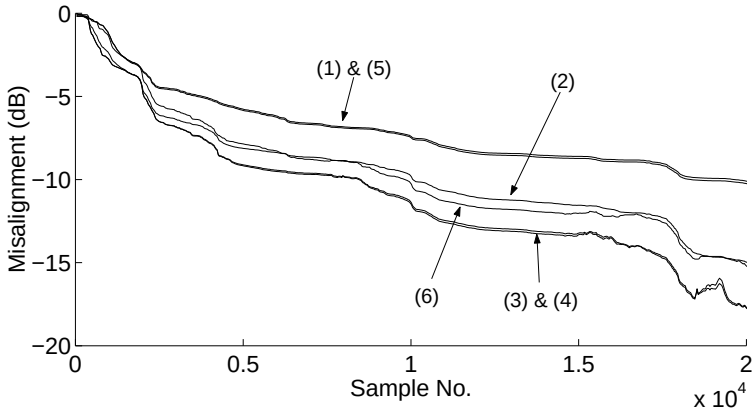

Figure 3: Misalignment Performance Comparison (Speech Input No. 1)— ϵ-NLMS Algorithm (SNR=30dB,L=50) — (1) ϵ=1.0, (2) ϵ=0.1, (3) ϵ=0.01, (4) ϵ: identical to the averaged variance of the channel-inputs, (5) $\epsilon=1/\mu$, (6) ϵ=0 with Channel-Input Power Consideration

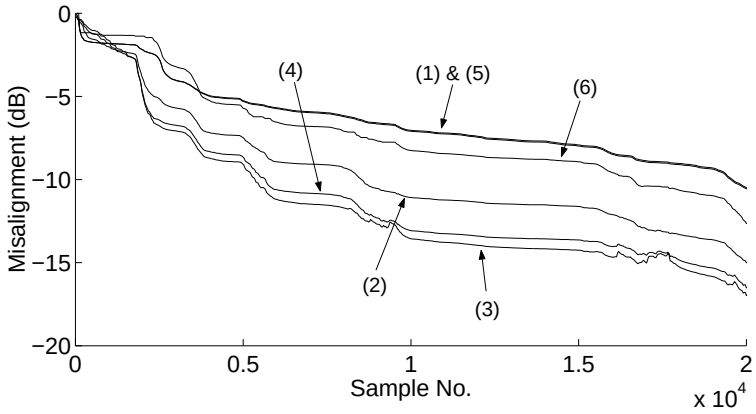

Figure 4: Misalignment Performance Comparison (Speech Input No. 2)— ϵ-NLMS Algorithm (SNR=30dB,L=50) — (1) ϵ=1.0, (2) ϵ=0.1, (3) ϵ=0.01, (4) ϵ: identical to the averaged variance of the channel-inputs, (5) $\epsilon=1/\mu$, (6) ϵ=0 with Channel-Input Power Consideration

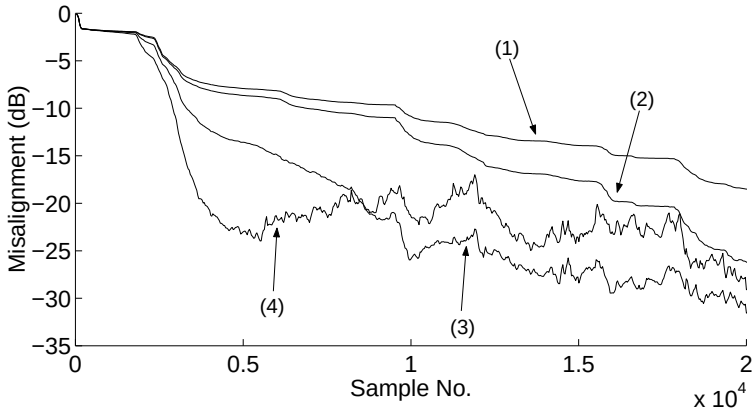

Figure 5: Misalignment Performance Comparison — FLS Algorithm (SNR=30dB,L=50) — (1) γ=0.99999, (2) γ=0.9999, (3) γ=0.9995, (4) γ=0.999

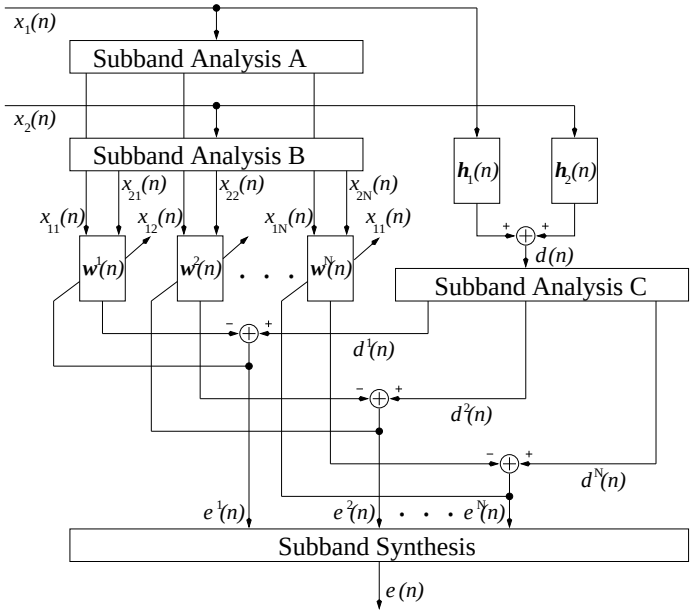

Figure 6: Structure of Subband Stereo Acoustic Echo Cancellation Scheme

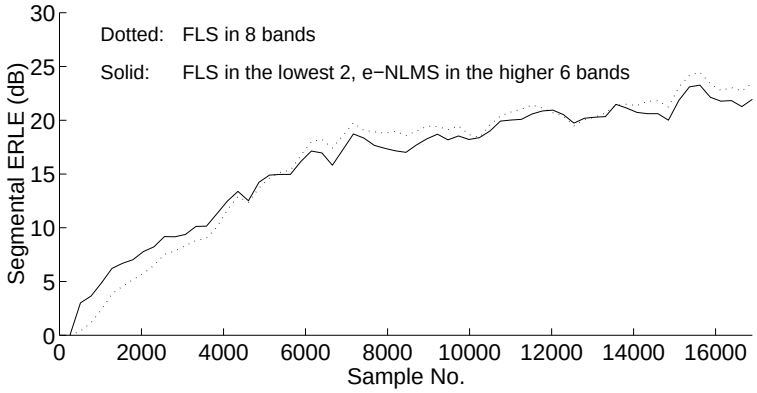

Figure 7: Comparison of the Segmental ERLE Performance (averaged over 24 different speech samples)

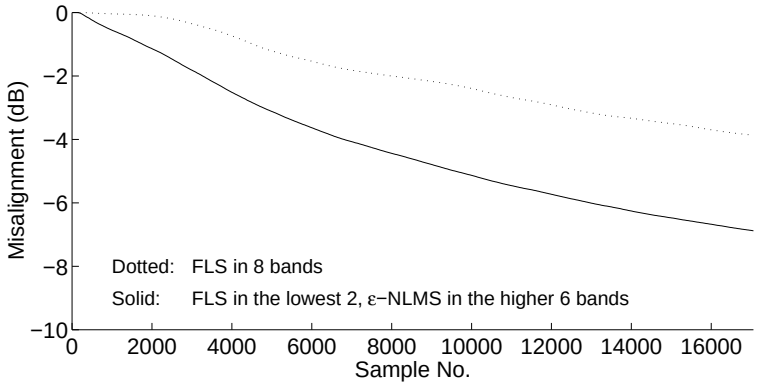

Figure 8: Comparison of the Misalignment Performance (averaged over 24 different speech samples)