LOW COMPLEXITY ϵ-NLMS ALGORITHMS AND SUBBAND STRUCTURES FOR STEREOPHONIC ACOUSTIC ECHO CANCELLATION

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

Communications and Signal Processing Group,
Department 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

One of the fundamental problems in Stereophonic Acoustic Echo Cancellation (SAEC) lies in the misadjustment in the filter coefficients due to the two strongly correlated channel-inputs. In this paper, we study the effect of the normalisation factor, ϵ, within the two-channel NLMS (ϵ-NLMS) algorithm for subband SAEC and show that the optimal choice may be close to the variance of the channel-input data. In the simulation study with real speech datasets, it is observed that subband stereo echo cancellers using the Fast Least Squares (FLS) algorithm in lower bands and the ϵ-NLMS algorithm, with the optimal normalisation factor setting, in higher frequency bands can improve misalignment performance significantly compared with using only the FLS algorithm in all subbands. In addition, around 2-3dB further misalignment performance improvement is obtained by applying smoothly time-varying allpass filters in the lower frequency bands, while introducing no perceptible auditory degradation.

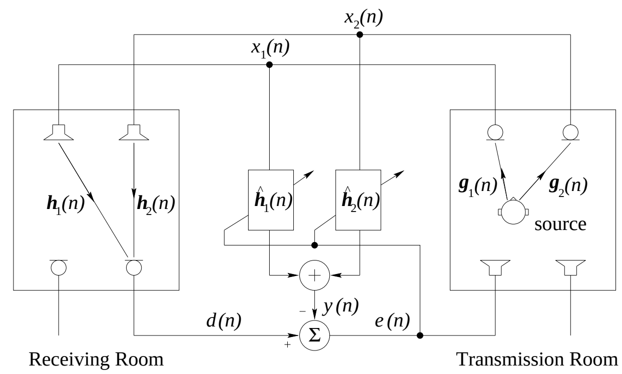


Figure 1: Schematic Diagram of Stereophonic Acoustic Echo Cancellation (only a single-channel echo-canceller is depicted for clarity)

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 diagram of Stereophonic Acoustic Echo Cancellation (SAEC).

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 takes into acount 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 numerical difficulty.

In the single channel case, it has been theoretically 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]. 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 Gaussian input data are provided. For the two-channel case such a non-linearity may whiten the original channel-inputs and thereby help de-correlate the two channel-inputs.

In this paper, we firstly investigate the effect of the normalisation factor in full-band SAEC through simulation study with real speech datasets (not white Gaussian input as used in [4, 5]) 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 [7, 8], 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 both the FLS algorithm in lower bands and the optmally-configured ϵ-NLMS algorithm in higher frequency bands gives significant improvement in terms of misalignment performance, while maintaining low computational complexity. In addition, a new smoothed version of the time-varying allpass filters in [9] are applied in the lower frequency (below 1kHz) bands in order to de-correlate the two channel-inputs. We then show that misalignment performance can thereby be improved further, while introducing much less auditory degradation, in comparison with the original full-band allpass filtering approach.

The contribution of this paper is therefore two-fold: (i) the study of the optimal normalisation factor settings within the two-channel ϵ-NLMS algorithm in subband SAEC (ii) the utility of the smoothed version of the time-varying allpass filters applied in only lower frequency bands to de-correlate further the two-channel inputs, while introducing no perceptible auditory degradation.

2. THE TWO-CHANNEL ϵ-NLMS TYPE ALGORITHMS

The update equation for the filter coefficients $\boldsymbol{h_1}$ and $\boldsymbol{h_2}$ at time index $k+1$ with the two-channel NLMS type algorithms is written in the form:

$$\begin{bmatrix} \boldsymbol{h}_1(k+1) \\ \boldsymbol{h}_2(k+1) \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_1(k) \\ \boldsymbol{h}_2(k) \end{bmatrix} + e(k) \begin{bmatrix} f_1(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) \\ f_2(\boldsymbol{x}_1(k), \boldsymbol{x}_2(k)) \end{bmatrix}, \quad (1)$$

where $e(k)$ is the error between the desired response $d(k)$ and sum value of the filter outputs:

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

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

2.1. The Two-Channel ϵ-NLMS Algorithm

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

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

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

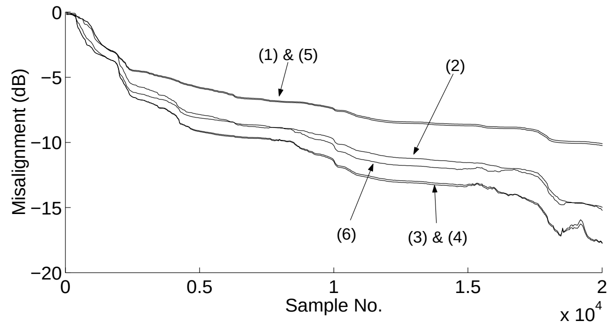
2.2. Experimental Evaluation for the Optimal Normalisation Setting

The ϵ-NLMS algorithm is initially applied in full-band SAEC. In the experimental evaluation, ϵ_i $(i = 1, 2)$ were kept identical for each channel, i.e., $\epsilon_i = \epsilon$. Firstly, ϵ was set to 1, 0.1, 0.01, 0 (with considering the channel-input power), and identical to the averaged *a priori* variance of the channel-input power. For the setting $\epsilon = 0$ with the power consideration, no weight update is 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 following modification in [5] for eqn. (3) was also used:

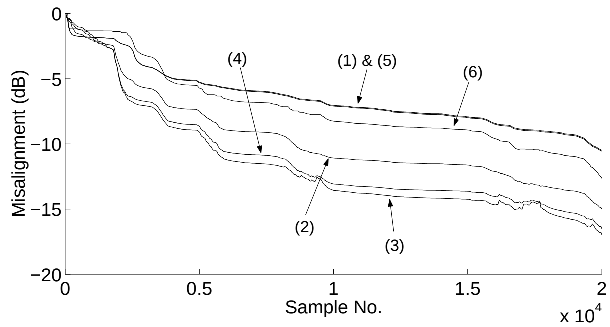
$$f'_{\mathrm{NLMS},i}(k) = \frac{1}{1/\mu + ||\boldsymbol{x}_1(k)||_2 + ||\boldsymbol{x}_2(k)||_2} \quad (i = 1, 2).$$

The channel-inputs and the modelled receiving room impulse responses used are the same as described in Section 4.

Figs. 2 (a) and (b) show the misalignment performance obtained for two different speech signals. Various values of ϵ are compared (including ϵ=$1/\mu$ as in [5]) in the ϵ-NLMS algorithm using L=50. The modelled receiving room impulse responses used were generated and modulated with variations simulated by a "random-walk" regression model [11]. In the figures, the simulation results are shown in the presence of noise in the echo-path in the receiving room at SNR=30dB.



(a) Speech Input No.1



(b) Speech Input No.2

Figure 2: Misalignment Performance Comparison — 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 *a priori* variance of the channel-inputs, (5) ϵ=$1/\mu$, (6) ϵ=0 with Channel-Input Power Consideration

2.3. Discussion

As shown in Fig. 2, 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 is stated a loose bound on the optimal range of ϵ 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 entire signal. In particular, this finding has been confirmed 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 estimates of $||\boldsymbol{x_1}||_2$ and $||\boldsymbol{x_2}||_2$ govern the adaptation gain; the presence of ϵ may well then be less effective.

In the evaluation study, it was also observed that the variations between the ϵ-NLMS algorithm with different normalisation factor settings used and the FLS algorithm becomes more apparent as the filter length increases. This indicates that the effective condition number of the input correlation matrix becomes much more significant as the number of the taps is increased.

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 numerical instability as well as the increase in computational complexity will be problematic, especially when the variance of the channel-input is very small (in fact, during the simulation, it was observed that the performance of the FLS algorithm was quite dependent upon the forgetting factor).

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

3. SUBBAND STEREO ECHO CANCELLER

3.1. Smoothly Time-Varying Allpass Filters

In [7, 10] a signal conditioning method based upon applying non-linearity in the channel-inputs is proposed in order to de-correlate the channel inputs, and its substantial de-correlation effect is demonstrated. In [9], another signal conditioning method based upon psychoacoutical studies and using time-varying allpass filters is proposed, and the method is applied in a full-band scheme. However, informal listening tests confirmed that the sound distortion is still heard with the parameter setting presented in that paper. In this paper, we use instead smoothly time-varying allpass filters in only lower frequency bands in order to maintain negligible auditory degradation. This is based upon the same principle proposed in [7]; when non-linear transformation is applied only in the lower frequency bands, the distortion is confined to the low-frequency band. The allpass filters used are the first-order filters described by a single parameter $\alpha_i(k)$ $(i = 1, 2)$ with the frequency response [9]

$$A_i(\omega, k) = \frac{e^{-j\omega} - \alpha_i(k)}{1 - \alpha_i(k)e^{-j\omega}} \tag{4}$$

In order for the stability of the allpass filters given above, the absolute values of $\alpha_i(k)$ at time instant k must be less than unity and also be real, as all signals are real. The update rule for $\alpha_i(k)$ is then given by

$$\begin{aligned} \alpha_i(k+1) &= \alpha_i(k) + r_i(k), \\ \text{set } \alpha_i(k+1) &= \alpha_{i,max} \text{ if } \alpha_i(k+1) > \alpha_{i,max}, \\ \text{set } \alpha_i(k+1) &= \alpha_{i,min} \text{ if } \alpha_i(k+1) < \alpha_{i,min}, \end{aligned} \tag{5}$$

where $\alpha_{i,max}$ and $\alpha_{i,min}$ are respectively set to 0 and -0.9 [9] chosen based upon psychoacoustical studies. $r_i(k)$ is an averaged value over a given window length $L_W = 30$ of an i.i.d. random variable having a uniform p.d.f. over the interval $[-0.3\ 0.3]$.

3.2. Structure

The structure of the subband stereo echo canceller used in this paper is depicted in Fig. 3. In the figure, $\hat{\boldsymbol{h}}^{\boldsymbol{i}}$ denotes the estimated filter coefficients vector in the ith subband and $\boldsymbol{h}^{\boldsymbol{i}}$ the vector of the modelling room impulse response.

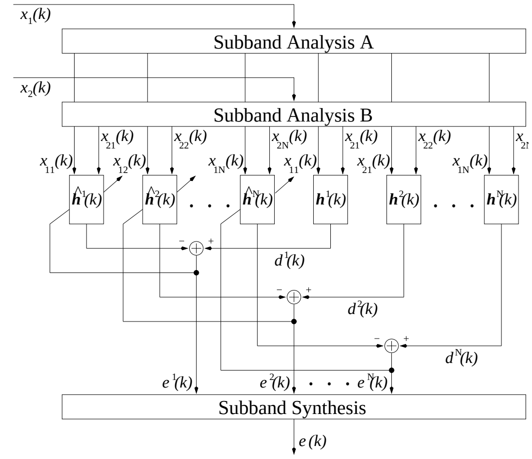


Figure 3: Subband Stereo Echo Canceller

4. SIMULATION STUDY

For the simulation study for subband SAEC, the lengths of the modelled receiving room impulse responses $\boldsymbol{h_1}$ and $\boldsymbol{h_2}$ were, in contrast, assumed to be $L = 1024$ and generated in the same way as in Section 2.2. In the simulation study, noise at SNR=30dB is assumed to be present in the echo-path in the receiving room. The channel-input signals are divided into eight $(N = 8)$ frequency subbands using a three-staged binary tree structure of halfband systems. The input data used were the real speech data recorded by two (i.e., a female and a male) different speakers in a quiet room, sampled originally at 48kHz, and down-sampled to 8kHz. The signals are divided into eight $(N = 8)$ frequency subbands using a three-staged binary tree structure of halfband systems. The filterbanks employed are FIR QMFs after Johnston [12].

The segmental Echo Loss Return Enhancement (ERLE) at scaled time instant j is given by

$$\text{ERLE}(j) = 10\log_{10} \sum_{i=1}^{256} \frac{d^2(256j + i)}{e^2(256j + i)}. \tag{6}$$

The misalignment performance is evaluated based upon the Averaged Weight Error Norm (A.W.E.N.) between the filter coefficient vectors in the jth frequency band $\boldsymbol{h}_1^j$ and $\boldsymbol{h}_2^j$ $(j = 1, 2, ..., N)$ and the optimum $\boldsymbol{h}_{i,opt}^j$:

$$\text{A.W.E.N.} = 10\log_{10} \frac{1}{2N} \sum_{i=1}^{2} \sum_{j=1}^{N} \frac{||\boldsymbol{h}_i^j - \boldsymbol{h}_{i,opt}^j||_2}{||\boldsymbol{h}_{i,opt}^j||_2}, \tag{7}$$

Figs. 4 and 6 show comparisons of the segmental ERLE performance. The misalignment performance obtained is shown in Figs. 5 and 7. The performance is compared between a subband stereophonic acoustic echo canceller using the FLS algorithm in all the subbands, a stereo echo canceller 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 without and with the smoothly time-varying allpass filters. For these simulations, the forgetting factor of the FLS algorithm was identically set to 0.999 in each subband.

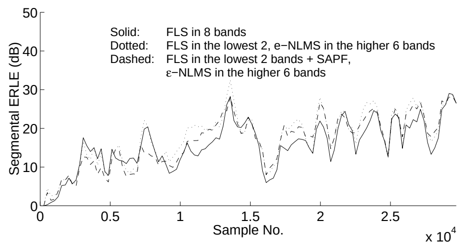


Figure 4: Comparison of the Segmental ERLE Performance, Speech Input No.1

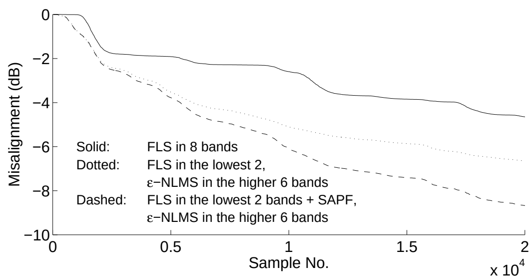


Figure 5: Comparison of the Misalignment Performance, Speech Input No.1

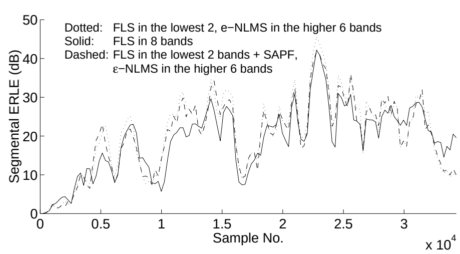


Figure 6: Comparison of the Segmental ERLE Performance, Speech Input No.2

5. CONCLUSION

In this paper, the utility of the two-channel ϵ-NLMS algorithm for SAEC has been studied 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 known *a priori* variance of the channel-inputs. In the simulation study, the results obtained by a subband stereo echo canceller using (i) the FLS algorithm with signal conditioning

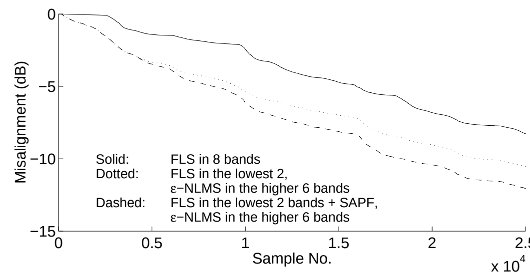


Figure 7: Comparison of the Misalignment Performance, Speech Input No.2

by the smoothed version of the time-varying allpass filters in the lower bands and (ii) the optimally configured ϵ-NLMS algorithm in the higher frequency bands show around 4-5dB misalignment performance improvement, while maintaining less computational complexity ($O(6 \times 4L + 2 \times 28L)$) than the echo canceller using the FLS algorithm in all the subbands ($O(8 \times 28L)$). Future work includes theoretical investigation of the optimally configured two channel ϵ-NLMS algorithm within subband SAEC.

6. REFERENCES

[1] M. M. Sondhi and 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, 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] 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.

[7] 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.

[8] 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.

[9] M. Ali, "Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-Pass Filtering For Signal Decorrelation", Proc. ICASSP-98, Vol. VI, pp. 3689-3692, Seattle, May 1998.

[10] 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, 1997.

[11] 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, Greece, Sept. 1998.

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