ROBUST ADAPTATION OF AN ADAPTIVE SIMO ECHO CANCELER

Tilman Huhn[2], *Hans-Joachim Jentschel*[1]

[1] Institut für Verkehrsinformationssysteme, Technische Universität Dresden
[2] Signal Processing Group, Zentrum Mikroelektronik Dresden

huhn@vini.vkw.tu-dresden.de

ABSTRACT

This paper refers to the multi-channel system identification of a linear single-input multiple-output (SIMO) system by an adaptive filter based on the normalised LMS algorithm. Configurations with an SIMO echocanceler are of special interest for the combination of echo cancelation and noise reduction principles. In this paper we investigate the problem of optimising a generalised adaptation parameter, called the stepsize matrix. The solution leads to a novel adaptation control approach, realising a coupling between the different channels. We will show that the channel coupling allows a partial cancellation of the measurement noise, which is adverse to the identification accuracy.

1. INTRODUCTION

Adaptive multiple-input multiple-output or multi-channel algorithms may be applied to signal processing in hands-free telephony, where a MC equipment is used, e.g. for stereophonic hands-free telephones [1,2] and for telephone sets with a combined noise and echo suppression device based on a microphone array [3,4], see Fig. 1.

Adaptive MC algorithms have been investigated in [2,5,6,7,8,9]. The main focus of these papers is directed to the convergence properties of the adaptive algorithms under the influence of cross-correlated loudspeaker signals, being the input of the unknown system. This problem and the proposed algorithms can be described in a convenient manner for the more special case of a multiple-input single-output (MISO) system.

A new idea is to consider, how the cross-correlation of the measurement noise influences on the convergence of the adaptation of a MC adaptive filter. We will exploit the results for a new MC NLMS algorithm. Without loss of generality we describe the principles for a single-input multiple-output (SIMO) system (one loudspeaker, several microphones). Such a configuration may be understood as a subconfiguration of a stereo-to-stereo hands-free equipment or as a subconfiguration of a combined echo and noise reduction system.

A coupled adaptation control of the adjacent single-channel filters is developed, based on a matrix notation of the MC system identification problem. In advantage to the conventional MC NLMS algorithm, popular for acoustic echo cancellation, the scalar stepsize parameter is replaced by a stepsize matrix.

The new adaptation control approach provides an improvement of the NLMS convergence properties for highly coherent measurement noise signals. Thus, the algorithm is especially useful, when the measurement noise is composed mainly from the near-end speech (i.e. in double-talk situations).

2. SYSTEM MODEL

For a closed notation of the SIMO system, matching the solid drawing in Fig. 1, we define the vectors $\mathbf{y}(k)=(y_1(k),\ldots,y_L(k))^T$ for the L-channel microphone signal, $\tilde{\mathbf{n}}(k)$ for the measurement noise and $\mathbf{d}(k)$ for the loudspeaker echo, respectively. The measurement noise $\tilde{n}_j(k)$ of each channel j, $j\in\{1, \ldots, L\}$, consists of two components: the environment noise $n_j(k)$ and the near-end speaker signal $s(k)$. Refering to the noise reduction approaches used in [10,3,4], it may be assumed, that the microphones are placed far enough from each other. Thus, both the diffuse background noise and the loudspeaker echo cause microphone signal components which are mutually uncorrelated. However, the speech signal is received from a distance lower than the reverberation radius and is assumed to be identical in all channels. The SIMO echo path is characterised by the composed impulse response vector $\mathbf{h}=(\mathbf{h}_1^T,\ldots,\mathbf{h}_L^T)^T$ with the single-channel impulse response vectors $\mathbf{h}_j=(h_j^T(0),\ldots,h_j^T(N-1))^T$. Let $\mathbf{x}(k)=(x(k),\ldots,x(k-N+1))^T$ be the vector of the input (loudspeaker) signal, the system equation can be written as

$$\mathbf{y}(k)=\mathbf{d}(k)+\tilde{\mathbf{n}}(k)=\mathbf{X}^T(k)\mathbf{h}(k)+\tilde{\mathbf{n}}(k), \tag{1}$$

where $\mathbf{X}(k)$ is the Kronecker product of $\mathbf{x}(k)$ with the $L\times L$ identity matrix $\mathbf{I}$:

$$\mathbf{X}(k)=\mathbf{I}\otimes\mathbf{x}(k)=\begin{pmatrix}\mathbf{x}(k) & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{x}(k) & & \mathbf{0}\\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{x}(k)\end{pmatrix} \tag{2}$$

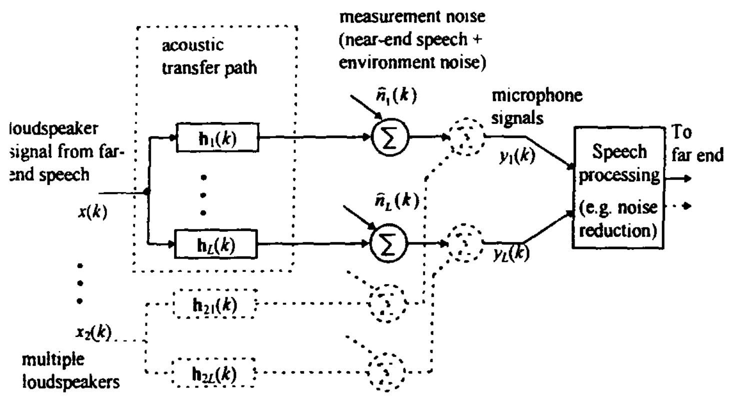

Fig. 1. Acoustic MC system (the SIMO subsystem is drawn by solid lines)

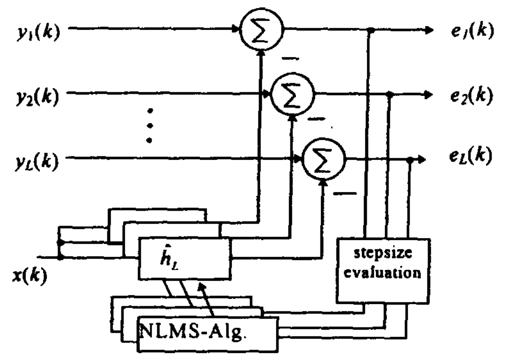

Fig. 2. SIMO NLMS algorithm

3. SIMO SYSTEM IDENTIFICATION

The formulation of the SIMO NLMS algorithm is straightforward from the conventional single-input single-output case:

$$\mathbf{e}(k) = \mathbf{y}(k) - \mathbf{X}^T(k)\hat{\mathbf{h}}(k) \tag{3}$$

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \frac{1}{\|\mathbf{x}(k)\|^2}\mathbf{X}(k)\mathbf{M}(k)\mathbf{e}(k) \tag{4}$$

There, $\mathbf{e}(k)$ is L-channel vector of the adaptation error signal and $\mathbf{M}(k)$ is a stepsize matrix controlling the adaptation speed of the algorithm. The vector $\hat{\mathbf{h}}(k)$ contains the coefficients of the adaptive transversal filter matching idealy the samples of the systems impulse response. In the conventional NLMS algorithm, used so far in MC indentification, the single channels are being adapted each one independently, no matter the cross-correlation of the microphone signals [11]. This can be expressed by writing the stepsize matrix $\mathbf{M}(k)$ in (4) as a diagonal matrix of the scalar single-channel stepsize parameters (Fig. 2):

$$\mathbf{M}_d(k) = \mathrm{diag}(\mu_1(k),\ldots,\mu_L(k)). \tag{5}$$

We call that uncoupled or conventional adaptation. If the stepsize parameters $\mu_j(k)$ are each optimised to minimise the expected evolution of convergence $\eta(k)$, described by

$$\eta(k) = \|\hat{\mathbf{h}}_\Delta(k)\|^2 - \|\hat{\mathbf{h}}_\Delta(k+1)\|^2, \tag{6}$$

their optimum arises from [13] to be

$$\mu_j(k) = \frac{\mathrm{E}\{r_j^2(k)\}}{\mathrm{E}\{e_j^2(k)\}}. \tag{7}$$

There, E{...} is the expectation operator, $\hat{\mathbf{h}}_\Delta(k)$ with

$$\hat{\mathbf{h}}_\Delta(k) = \mathbf{h}(k) - \hat{\mathbf{h}}(k) \tag{8}$$

is the parameter error. The residual echo r describes that part of the echo that can not be cancelled using the current parameter vector $\hat{\mathbf{h}}(k)$. The diagonal stepsize matrix (5) with (7) will be refered to as the optimum diagonal stepsize matrix $\mathbf{M}_d^*(k)$.

A novel adaptation approach is obtained by cross-coupling of the rows of vector $\mathbf{e}(k)$. We can accomplish this by allowing all components of $\mathbf{M}(k)$ to be nonzero. We will term this more general choice of $\mathbf{M}(k)$ as the coupled adaptation. The minimum mean coefficient error norm will be derived in the appendix to be:

$$\mathbf{M}^*(k) = (\mathbf{R}_{ee}(k))^{-1}\mathbf{R}_{rr}(k), \tag{9}$$

where $\mathbf{R}_{rr}(k)$ and $\mathbf{R}_{ee}(k)$ are the MC covariance matrices of the residual echo and the error signal noise, respectively. For the latter the following identity holds:

$$\mathbf{R}_{ee}(k) = \mathbf{R}_{rr}(k) + \mathbf{R}_{\hat{n}\hat{n}}(k), \tag{10}$$

where $\mathbf{R}_{\hat{n}\hat{n}}(k) = \mathrm{E}\{\hat{\mathbf{n}}(k)\hat{\mathbf{n}}^T(k)\}$ is the covariance matrix of the measurement noise.

4. COMPARISON OF COUPLED AND UNCOUPLED ADAPTATION

The expected evolution of convergence $\mathrm{E}\{\eta(k)\}$ will be compared for both of the optimal stepsize matrices (7) and (9). An analytic expression can be derived for the stepsize matrices, if we consider the measurement conditions, described by the assumptions on the mutual channel-coherence in paragraph 2. Under these assumptions the signal s corresponds with the coherent component of the measurement noise. Thus, the covariance matrices can be stated as

$$\mathbf{R}_{\hat{n}\hat{n}}(k) = \sigma_n^2(k)\,\mathbf{I} + \sigma_s^2(k)\,\mathbf{1}\mathbf{1}^T$$

$$\mathbf{R}_{rr}(k) = \sigma_r^2(k)\,\mathbf{I}\,,$$

where σ_n^2, σ_s^2 and σ_r^2 are the short-time variances of the corresponding signals in each of the L channels. The symbol $\mathbf{1}$ stands for the 1-filled column vector of dimension L.

After some algebra [12] we get the optimum stepsize matrices

$$\mathbf{M}^*(k) = \mu_{\mathrm{ST}}(k)\cdot(\mathbf{I} - w(k)\cdot\mathbf{1}\mathbf{1}^T) \text{ and} \tag{11}$$

$$\mathbf{M}_d^*(k) = \mu_d(k)\cdot\mathbf{I}. \tag{12}$$

The symbols $\mu_{\mathrm{ST}}(k)$, $w(k)$ and $\mu_d(k)$ stand for the terms

$$\mu_{ST}(k) = \frac{\sigma_r^2(k)}{\sigma_r^2(k) + \sigma_n^2(k)} \quad (13)$$

$$w(k) = \frac{\sigma_s^2(k)}{\sigma_r^2(k) + \sigma_n^2(k) + L\sigma_s^2(k)} \quad (14)$$

$$\mu_d(k) = \frac{\sigma_r^2(k)}{\sigma_r^2(k) + \sigma_n^2(k) + \sigma_s^2(k)}. \quad (15)$$

The stepsize $\mu_d(k)$ simply reflects the stepsize of (7) for the assumed conditions. However, the coupled stepsize matrix in (11) requires some more attention. It consists of two factors influencing the convergence behaviour. The stepsize $\mu_{ST}(k)$ may be interpreted as the optimal single-talk stepsize, since the term $w(k)\mathbf{1}\mathbf{1}^T$ vanishes when the near-end speaker is silent. In contrast, the term $(\mathbf{I}-w(k)\mathbf{1}\mathbf{1}^T)$ applied to the error vector $\mathbf{e}(k)$ in (2) and (3) describes the cancellation of the coherent measurement noise component. For L=1, both stepsize matrices in (11) and (12) are identical, as it should be expected.

The expected evolution of convergence at the adaptation start k=0 may be directly derived from (11) and (12) [12]. A normalised term of $E\{\eta(0)\}$ is plotted in Fig. 3 versus the number L of channels for a sample setting of the short-time variances. Obviously, the convergence behaviour of the coupled adaptation approach improves as the number of channels increases. In limit it tends to the value, which would be reached in the single-talk situation $\sigma_s^2 = 0$. Alternatively spoken, the better some information about the coherent component s of the measurement noise $\tilde{n}$ is extracted from the signals, the better $\tilde{n}$ can be canceled in the error vector by $\mathbf{M}^*(k)\mathbf{e}(k)$.

The mean square coherence of the measurement noise is $\gamma^2(k) = \sigma_s^2(k)/\sigma_{\tilde{n}}^2(k)$. It corresponds with the microphone input SNR, which is defined by

$$SNR = 10dB \cdot \log_{10}\left(\sigma_s^2(k)/\sigma_n^2(k)\right). \quad (16)$$

We simulated both of the adaptation approaches under different conditions with N=200 for a simulation time of 20000 samples and recorded the relative parameter error norm:

$$P(k) = 10dB \cdot \log_{10}\left(\left\|\hat{\mathbf{h}}_\Delta(k)\right\|^2 / \left\|\mathbf{h}(k+1)\right\|^2\right). \quad (17)$$

This measure is approximately identical to the negative echo return loss enhancement, offen used in echo cancellation literature. Fig. 4 shows the relative parameter error norm versus the quadratic coherence $\gamma^2(k)$ of the measurement noise at the time instant k=20000. For better interpretation, the SNR ordinate is added to the plot. Two basic signal situations have been simulated. In Fig. 4a), white noise signals have been used. Fig. 4b) is based on realistic speech and environment signals. It is two be seen in both cases that the coupled adaptation provides faster convergence than the uncoupled one.

5. CONCLUSIONS

The paper presents a novel approach for the adaptation of parallel working adaptive filters, applied all to the same input signal and to different output signals, originating from the same source signals. When the system output signal contains a measurement noise independend from the input signal, then the optimisation of the above generalised stepsize matrix leads to a coupled adaptation of the single adaptive filters. It have been shown analytically and by simulation, that the coupled adaptation gains in better convergence behaviour than the conventional uncoupled adaptation. The next challenging question is the practical adjustment of the stepsize matrix by means of esimated covariance information. One possible way is to feed back the output signal of a noise reduction system processing at the sum error signal [3].

The results of this paper are of special interest for the application to SIMO echo cancelers in microphone arrays for combined echo and noise reduction.

ACKNOWLEDGEMENT

This paper is related to the project "Design and implementation of combined adaptive systems for hands-free telephony" supported by the Centre of Technology of Telekom, TZ Darmstadt/ Berlin.

APPENDIX

Derivation of the optimal Stepsize Matrix

From (6), (8) and (3) we get

$$\eta(k) = \left\|\hat{\mathbf{h}}_\Delta(k)\right\|^2 - \Bigg(\left\|\hat{\mathbf{h}}_\Delta(k)\right\|^2 - 2\frac{1}{\left\|\mathbf{x}(k)\right\|^2}\hat{\mathbf{h}}_\Delta^T(k)\mathbf{X}(k)\mathbf{M}(k)\mathbf{e}(k) + \frac{1}{\left(\left\|\mathbf{x}(k)\right\|^2\right)^2}\mathbf{e}^T(k)\mathbf{M}^T(k)\mathbf{X}^T(k)\mathbf{X}(k)\mathbf{M}(k)\mathbf{e}(k)\Bigg).$$

We simplify this using the relation $\mathbf{X}^T(k)\mathbf{X}(k) = \left\|\mathbf{x}(k)\right\|^2\mathbf{I}_L$.:

$$\eta(k) = \frac{1}{\left\|\mathbf{x}(k)\right\|^2}\Big(2\mathbf{r}^T(k)\mathbf{M}(k)\mathbf{e}(k) - \mathbf{e}^T(k)\mathbf{M}^T(k)\mathbf{M}(k)\mathbf{e}(k)\Big).$$

The expectation leads to

$$E\{\eta(k)\} = E\left\{\frac{1}{\left\|\mathbf{x}(k)\right\|^2}\Big(2\mathbf{r}^T(k)\mathbf{M}(k)\mathbf{e}(k) - \mathbf{e}^T(k)\mathbf{M}^T(k)\mathbf{M}(k)\mathbf{e}(k)\Big)\right\}.$$

We further state two often used assumptions [13] – the term $\left\|\mathbf{x}(k)\right\|^2$ will be seen as time-indepentend constant and the terms $E\{r_j e_i\}$ can be substituted by $E\{r_j r_i\}$, due to the statistical independence and the vanishing correlation between measurement noise $\hat{n}$ and residual echo r_j. Thus, the final result about the evolution of convergence as a function of the stepsize matrix **M** is:

$$E\{\eta(k)\} = \frac{1}{\|\mathbf{x}(k)\|^2}\Big(2\,E\big\{\mathbf{r}^T(k)\mathbf{M}(k)\mathbf{r}(k)\big\} - E\big\{\mathbf{e}^T(k)\mathbf{M}^T(k)\mathbf{M}(k)\mathbf{e}(k)\big\}\Big)$$

By differentiation of this term with respect to the stepsize matrix and setting it to zero, we get a relation for the maximum expected evolution of convergence and finally the corresponding stepsize matrix in (9). For the differentation with respect to a matrix see [12].

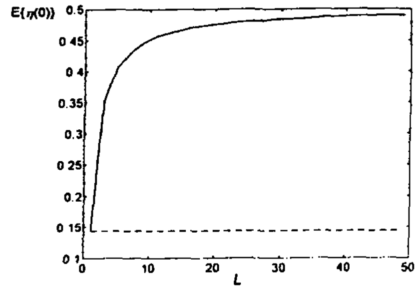

Fig. 3. Expected evolution of convergence at time instant k=0 for optimal coupled —— and uncoupled adaptation - - - - (with parameters $\sigma_s^2 = \sigma_n^2 = \sigma_r^2(k) = 1$)

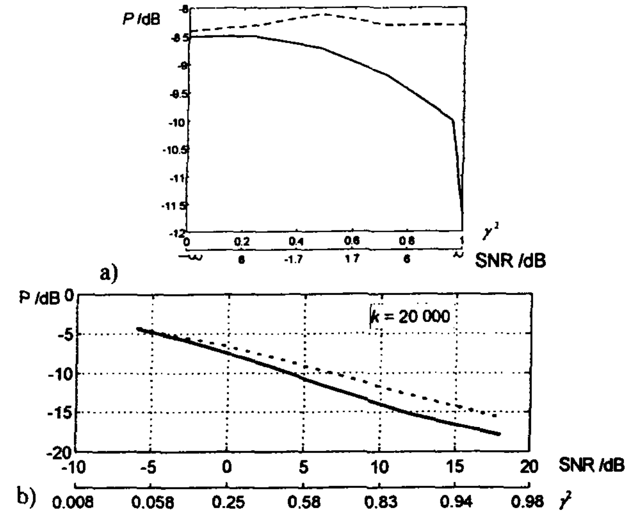

Fig. 4. Relative coefficient error norm versus varying quadratic coherence of the near-end measurement for coupled —— and conventional - - - - adaptation control at one of the microphone channels; a) for white noise signals b) for realistic signals

REFERENCE

[1] Sondhi, M.M. ; Morgan, D.R.: Acoustic Echo Cancellation for Stereophonic Teleconferencing, *IEEE Workshop on Audio and Acoustics*, Mohonc (NY) 1991

[2] Benesty, J. ; Morgan, D. ; Sondhi, M.: A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation, *Intern. Conf. on Acoustics, Speech and Signal Processing.* ICASSP '97, Munich, 1997, Vol. I, pp.303-306

[3] Huhn,T. ; Jentschel, H.-J.: Combination of Noise Reduction and Echo Cancellation in Hands-free Telephony (in German), *Nachrichtentechnik, Elektronik*, Berlin-Munich, 43 (1993) 6, pp.274-280

[4] Martin, R. ; Vary, P.: Combined acoustic echo cancellation, dereverberation and noise reduction: a two microphone approach, *Annales des Telecommunications* 49(1994), 429-438.

[5] Hirano, A. ; Sugiyama, A.: A Compact Multi-Channel Echo Canceller with a Single Adaptive Filter per Channel, *Intern. Symp. on Circuits and Systems*, ISCAS '92, San Diego (Cal.), 1992, pp.1922-1925

[6] Shimauchi, S. ; Makino, S.: Stereo Projection Echo Canceller with True Echo Path Estimation, *Intern. Conf. on Acoustics, Speech and Signal Processing*, ICASSP '95, Detroit (Michigan), 1995, pp.3059-3062

[7] Shimauchi, S. ; Makino, S.: Stereophonic Echo Cancellation Algorithm using Imaginary Input-Output Relationships, *Intern. Conf. on Acoustics, Speech and Signal Processing*, ICASSP '96, Atlanta (Georgia), 1996, pp.941-944

[8] Amand, F. ; Benesty, J. ; Gilloire, A. ; Grenier, Y.: A Fast Two-Channel Projection Algorithm for Stereophonic Acoustic Echo Cancellation, *Internat.Conf. on Acoustic Speech and Signal Processing*, ICASSP '96, Atlanta (Georgia), 1995, pp.949-952

[9] Benesty, J. ; Amand, F. ; Gilloire, A. ; Grenier, Y.: Adaptive Filtering Algorithms for Stereophonic Acoustic Echo Cancellation, *Internat.Conf. on Acoustic Speech and Signal Processing*, ICASSP '95, Detroit (Michigan), 1995, pp.3099-3102

[10] Zelinski,R.: A Microphone Array with Adaptive Post-filtering for Noise Reduction in Reverberant Rooms, *ICASSP'88*, New York (NY.), 1988, pp.2578-2581

[11] Ljung, L.: *Systemidentification, Theory for the User*, Englewood Cliffs, N.J., Prentice-Hall, 1987

[12] Huhn, T. ; Jentschel, H.-J.: *Adaptive Algorithms for a multi-channel echo canceler* (in German), TU Dresden, Oktober 1996 – Report of a research project supported by the Centre of Technology of Deutsche Telekom, TZ Darmstadt/ Berlin

[13] Schultheiß, U.: *On the Cancellation of an Acoustic Echo Canceler* (in German), Ph.D., TH Darmstadt, 1988