Stereophonic Acoustic Echo Cancellation: Analysis of the Misalignment in the Frequency Domain

Andy W. H. Khong, Jacob Benesty, *Senior Member, IEEE*, and Patrick A. Naylor, *Member, IEEE*

***Abstract*—The performance in terms of misalignment of adaptive algorithms, in general, is dependent on the conditioning of the input signal covariance matrix. The performance of two-channel adaptive algorithms is further degraded by the high interchannel coherence between the two input signals. In this letter, we establish the relationship between interchannel coherence of the two input signals and condition of the corresponding covariance matrix for stereo acoustic echo cancellation application. We show how this relationship affects the misalignment of a frequency-domain adaptive algorithm. We provide simulation results for both white Gaussian noise and speech input to verify our mathematical analysis.**

***Index Terms*—Condition number, interchannel coherence, misalignment, stereophonic acoustic echo cancellation.**

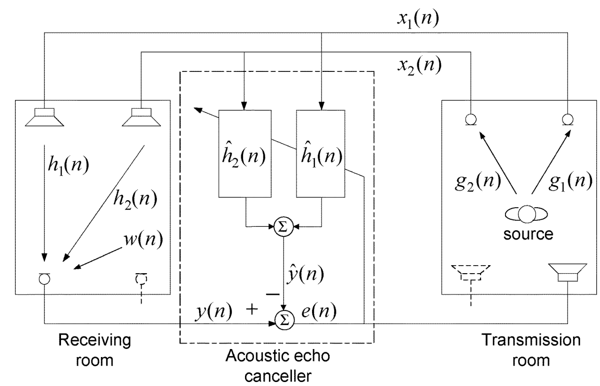

Fig. 1. SAEC system.

I. Introduction

In hands-free teleconferencing systems, multi-input and multi-output (MIMO) transmission provides telepresence by enhancing source localization. The stereophonic acoustic echo canceller (SAEC), such as shown in Fig. 1, suppresses the echo returned to the transmission room so as to enable undisturbed communication between the rooms.

A serious problem encountered in SAEC is the nonuniqueness problem [1]. It has been shown [2] that for a practical stereophonic system, the tap-input covariance matrix $\mathbf{R}$ is highly ill-conditioned. This is due to the high coherence between the two input signals $x_1(n)$ and $x_2(n)$, which in turn degrades the misalignment performance of adaptive algorithms. Many proposed solutions have since been introduced using, for example, nonlinear processing [1], [3], spectrally shaped random noise [4], time-varying all-pass filtering [5], and algorithms employing tap-selection [6]. The aim of these algorithms is to achieve decorrelation of $x_1(n)$ and $x_2(n)$, hence improving the condition of $\mathbf{R}$, without affecting the quality or stereophonic image of the speech. A survey of existing techniques for SAEC can be found in [7].

It has been shown [8] that for a single channel case, the performance of adaptive algorithms in terms of their final misalignment is affected by the conditioning of $\mathbf{R}$. For the stereo case, however, although it has generally been noted that the conditioning of $\mathbf{R}$ is degraded by the high interchannel coherence between $x_1(n)$ and $x_2(n)$, no explicit relationship between the two has been established. The aim of this letter is to establish this relationship that allows one to gain an insight of how interchannel coherence degrades the final misalignment performance of SAEC algorithms through the ill-conditioning of $\mathbf{R}$. Using this relationship, one can determine the level of ill-conditioning of $\mathbf{R}$ through the interchannel coherence estimate and design regularization parameters so as to improve the conditioning of $\mathbf{R}$, hence giving good misalignment performance, such as shown in [9]. We will verify the validity of the established relationship for both white Gaussian noise (WGN) and speech input signals by showing how the condition number is affected by the interchannel coherence, which in turn affects the performance of a two-channel frequency-domain adaptive algorithm [3] in terms of its final misalignment.

Manuscript received July 14, 2005; revised September 8, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dominic K. C. Ho.

A. W. H. Khong and P. A. Naylor are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: andy.khong@imperial.ac.uk; p.naylor@imperial.ac.uk).

J. Benesty is with the Université du Québec, INRS-EMT, Montréal, QC H5A 1K6, Canada (e-mail: benesty@emt.inrs.ca).

Digital Object Identifier 10.1109/LSP.2005.860545

II. Link Between Interchannel Coherence and Condition Number

We first express the covariance matrix $\mathbf{R}$ in terms of its auto- and cross-spectra content and then exploit the E-norm condition number [8], as will be described in this section. We denote $\mathbf{x}_j(n) = [x_j(n), \ldots, x_j(n-L+1)]^T$ as the tap-input vector for channels $j = 1, 2$ such that L and the superscript T are the length of the adaptive filter and vector transposition, respectively. Let $\mathbf{x}(n) = [\mathbf{x}_1^T(n)\mathbf{x}_2^T(n)]^T$, and the two-channel covariance matrix is thus given by

$$\mathbf{R} = E\left\{\mathbf{x}(n)\mathbf{x}^T(n)\right\} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix}_{2L \times 2L} \quad (1)$$

where $E\{\cdot\}$ is the mathematical expectation operator. It had been noted that adaptive algorithms aim to solve the Wiener–Hopf equations [10] given by

$$\widehat{\mathbf{h}} = \mathbf{R}^{-1}\mathbf{p} \quad (2)$$

where $\mathbf{p} = E\{\mathbf{x}(n)y(n)\}$ and $\widehat{\mathbf{h}} = [\widehat{\mathbf{h}}_1^T \; \widehat{\mathbf{h}}_2^T]^T$. It is evident from (2) that an ill-conditioned $\mathbf{R}$ will yield a bad estimate of $\widehat{\mathbf{h}}$. The performance of adaptive algorithms in SAEC is further degraded by the interchannel coherence between $x_1(n)$ and $x_2(n)$, as will be shown below.

A. Covariance Matrix and Spectra Content

We first note that for $L \to \infty$, a Toeplitz matrix is asymptotically equivalent to a circulant matrix if its elements are absolutely summable [11]. Defining $i = \sqrt{-1}$, we can decompose the $L \times L$ covariance matrix between the jth and kth channel $\mathbf{R}_{jk}$, given in (1) as [3]

$$\mathbf{R}_{jk} = \mathbf{F}^{-1}\mathbf{S}_{jk}\mathbf{F}, \qquad j,k = 1,2 \tag{3}$$

where $\mathbf{F}$ is the Fourier matrix defined with elements $F_{pq} = e^{-i2\pi pq/L}$ for $p, q = 0, 1, \ldots, L-1$. The $L \times L$ matrix

$$\mathbf{S}_{jk} = \mathrm{diag}\left\{S_{jk}(0), S_{jk}(1), \ldots, S_{jk}(L-1)\right\} \tag{4}$$

contains elements corresponding to the L frequency bins, which are formed from the discrete Fourier transform (DFT) of the first column of $\mathbf{R}_{jk}$. Letting $r_{jk}(l)$ be the auto- and cross-correlation coefficients for $j = k$ and $j \neq k$, respectively, we may now see that the spectra content between two signals is related to the correlation function by

$$S_{jk}(f) = \sum_{l=-\infty}^{\infty} r_{jk}(l)e^{-i2\pi fl}, \qquad f = 0, 1, \ldots, L-1. \tag{5}$$

Using (3), $\mathbf{R}$ can be expressed in terms of its spectra as

$$\mathbf{R} = \begin{bmatrix} \mathbf{F}^{-1} & \mathbf{0}_{L\times L} \\ \mathbf{0}_{L\times L} & \mathbf{F}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{F} & \mathbf{0}_{L\times L} \\ \mathbf{0}_{L\times L} & \mathbf{F} \end{bmatrix} \tag{6}$$

where $\mathbf{0}_{L\times L}$ is a null matrix of dimension $L \times L$.

B. E-Norm Condition Number of Covariance Matrix

The condition number $\chi[\mathbf{A}]$ of any $2L \times 2L$ matrix $\mathbf{A}$ is commonly computed using the l_2-norm [12] and is denoted by $\chi_2[\mathbf{A}] = \|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2 = \overline{\lambda}_{2L-1}/\overline{\lambda}_0$, where $\|\cdot\|_2$ is the l_2-norm operator, and $\overline{\lambda}_q$ is the qth eigenvalue of $\mathbf{A}$ such that $0 \leq \overline{\lambda}_0 \leq \overline{\lambda}_1 \leq, \ldots, \leq \overline{\lambda}_{2L-1}$. However, it has been shown [8] that the E-norm is suitable for our application, as we will explain briefly in this section. We first note that the symmetric and positive definite covariance matrix can be diagonalized as $\mathbf{Q}^T\mathbf{R}\mathbf{Q} = \Lambda$, where $\mathbf{Q}$ is a unitary matrix such that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ and $\Lambda = \mathrm{diag}\{\overline{\lambda}_0, \overline{\lambda}_1, \cdots, \overline{\lambda}_{2L-1}\}$ containing the eigenvalues of $\mathbf{R}$ with $0 \leq \overline{\lambda}_0 \leq \overline{\lambda}_1 \leq, \ldots, \leq \overline{\lambda}_{2L-1}$. By definition

$$\mathbf{R}^{\frac{1}{2}} = \mathbf{Q}\Lambda^{\frac{1}{2}}\mathbf{Q}^T. \tag{7}$$

Defining $\mathrm{tr}\{\cdot\}$ as the trace operator, the E-norm of the $2L \times 2L$ matrix $\mathbf{R}$ is then defined as [8]

$$\|\mathbf{R}\|_{\mathrm{E}} = \left[\frac{1}{2L}\mathrm{tr}\{\mathbf{R}^T\mathbf{R}\}\right]^{\frac{1}{2}}. \tag{8}$$

Noting that the squared Frobenius-norm [12] of $\mathbf{R}$ is defined as $\|\mathbf{R}\|_{\mathrm{F}}^2 = \mathrm{tr}\{\mathbf{R}^T\mathbf{R}\}$, the E-norm is then equivalent to the F-norm scaled by a factor $1/\sqrt{2L}$. Using (7), it follows that

$$\left\|\mathbf{R}^{\frac{1}{2}}\right\|_{\mathrm{E}} = \left[\frac{1}{2L}\mathrm{tr}\{\mathbf{R}\}\right]^{\frac{1}{2}} \tag{9}$$

$$\left\|\mathbf{R}^{-\frac{1}{2}}\right\|_{\mathrm{E}} = \left[\frac{1}{2L}\mathrm{tr}\{\mathbf{R}^{-1}\}\right]^{\frac{1}{2}} \tag{10}$$

which results in the E-norm condition number

$$\chi_{\mathrm{E}}\left[\mathbf{R}^{\frac{1}{2}}\right] = \left\|\mathbf{R}^{\frac{1}{2}}\right\|_{\mathrm{E}}\left\|\mathbf{R}^{-\frac{1}{2}}\right\|_{\mathrm{E}}. \tag{11}$$

We may now see that the E-norm of the identity matrix is one, and if $\chi_{\mathrm{E}}[\mathbf{R}^{1/2}]$ is large, the covariance matrix $\mathbf{R}$ is said to be ill-conditioned. In addition, for our SAEC application, since we would like to study only the effect of interchannel coherence on the condition number, the factor 1/(2L) removes the dependency of condition number on L, hence making $\chi_{\mathrm{E}}[\mathbf{R}^{1/2}]$ a suitable measure. It is further shown in [8] that $\chi_{\mathrm{E}}^2[\mathbf{R}^{1/2}]$ is a good measure of the conditioning of $\mathbf{R}$.

C. Relationship Between Interchannel Coherence and Condition Number of R

To compute $\chi_{\mathrm{E}}[\mathbf{R}^{1/2}]$, we first define $\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}$ and start by computing $\mathrm{tr}\{\mathbf{R}\}$ from (6) using the following:

$$\mathrm{tr}\{\mathbf{R}\} = \mathrm{tr}\{\breve{\mathbf{F}}^{-1}\mathbf{S}\breve{\mathbf{F}}\} = \sum_{l=0}^{L-1}[S_{11}(l) + S_{22}(l)] \tag{12}$$

where we have defined the $2L \times 2L$ matrices

$$\breve{\mathbf{F}} = \begin{bmatrix} \mathbf{F} & \mathbf{0}_{L\times L} \\ \mathbf{0}_{L\times L} & \mathbf{F} \end{bmatrix} \quad \breve{\mathbf{F}}^{-1} = \begin{bmatrix} \mathbf{F}^{-1} & \mathbf{0}_{L\times L} \\ \mathbf{0}_{L\times L} & \mathbf{F}^{-1} \end{bmatrix} \tag{13}$$

and have used the relation $\mathrm{tr}\{\mathbf{AB}\} = \mathrm{tr}\{\mathbf{BA}\}$. Using (13), we now compute $\mathrm{tr}\{\mathbf{R}^{-1}\}$, giving

$$\mathrm{tr}\{\mathbf{R}^{-1}\} = \mathrm{tr}\{\breve{\mathbf{F}}^{-1}\mathbf{S}^{-1}\breve{\mathbf{F}}\} = \mathrm{tr}\{\mathbf{S}^{-1}\}. \tag{14}$$

We first note that for the trivial case of $x_1(n) = x_2(n)$, $\mathbf{R}^{-1}$ does not exist. Using a similar approach to [3] and [13], provided that the coherence is not equal to one for any frequency, we then express the $2L \times 2L$ inverse matrix as

$$\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{S}_1^{-1} & \mathbf{0}_{L\times L} \\ \mathbf{0}_{L\times L} & \mathbf{S}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{L\times L} & -\mathbf{S}_{12}\mathbf{S}_{22}^{-1} \\ -\mathbf{S}_{21}\mathbf{S}_{11}^{-1} & \mathbf{I}_{L\times L} \end{bmatrix} \tag{15}$$

where $\mathbf{I}_{L\times L}$ is an $L \times L$ identity matrix and the submatrices

$$\mathbf{S}_1 = \left[\mathbf{I}_{L\times L} - \mathbf{S}_{12}^2\left(\mathbf{S}_{11}^{-1}\mathbf{S}_{22}^{-1}\right)\right]\mathbf{S}_{11} \tag{16}$$

$$\mathbf{S}_2 = \left[\mathbf{I}_{L\times L} - \mathbf{S}_{12}^2\left(\mathbf{S}_{11}^{-1}\mathbf{S}_{22}^{-1}\right)\right]\mathbf{S}_{22}. \tag{17}$$

From (5), we note that the squared interchannel coherence function of the fth frequency bin may be expressed in terms of the spectra of input signals as

$$|\gamma(f)|^2 = \frac{|S_{12}(f)|^2}{S_{11}(f)S_{22}(f)}, \qquad f = 0, 1, \ldots, L-1 \tag{18}$$

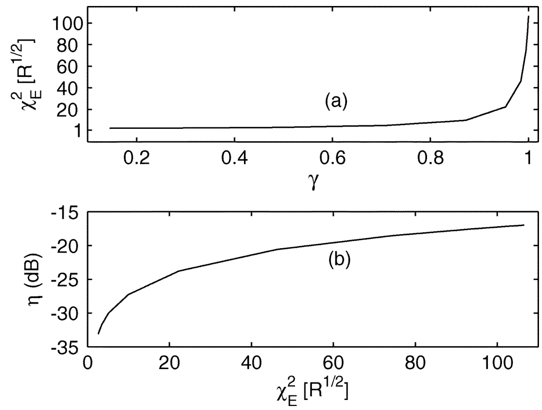

Fig. 2. Variation of (a) $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$ with γ and (b) η with $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$.

and we define the diagonal coherence matrix given by

$$|\Gamma|^2 = \mathrm{diag}\left\{|\gamma(0)|^2, |\gamma(1)|^2, \ldots, |\gamma(L-1)|^2\right\}. \tag{19}$$

We may now express the diagonal matrices $\mathbf{S}_1^{-1}$ and $\mathbf{S}_2^{-1}$ of (15) in terms of (19) as

$$\mathbf{S}_1^{-1} = \left[\mathbf{I}_{L\times L} - |\Gamma|^2\right]^{-1}\mathbf{S}_{11}^{-1} \tag{20}$$
$$\mathbf{S}_2^{-1} = \left[\mathbf{I}_{L\times L} - |\Gamma|^2\right]^{-1}\mathbf{S}_{22}^{-1} \tag{21}$$

and hence, we can now simplify (14), giving

$$\mathrm{tr}\{\mathbf{S}^{-1}\} = \sum_{l=0}^{L-1}\left[1 - |\gamma(l)|^2\right]^{-1}\left[S_{11}^{-1}(l) + S_{22}^{-1}(l)\right]. \tag{22}$$

Substituting (12) and (22) into (11), we finally obtain the relationship between interchannel coherence and E-norm condition number of $\mathbf{R}$, given as

$$\chi^2_{\mathrm{E}}\left[\mathbf{R}^{\frac{1}{2}}\right] = \frac{1}{4L^2}\left[\sum_{l=0}^{L-1}\left[S_{11}(l) + S_{22}(l)\right]\right] \times \left[\sum_{l=0}^{L-1}\left[1 - |\gamma(l)|^2\right]^{-1}\left[S_{11}^{-1}(l) + S_{22}^{-1}(l)\right]\right]. \tag{23}$$

We first note that the computation of $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$ is tractable since $\mathbf{S}_{11}$, $\mathbf{S}_{22}$, and Γ are diagonal matrices. More importantly, it is now evident from (23) that $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$ increases with the squared interchannel coherence function, hence degrading the condition of $\mathbf{R}$. Fig. 2(a) shows how $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$ varies with interchannel coherence γ for an example case of $L = 1024$ with unit variance WGN stereophonic input. Defining

$$\eta(n) = 10\log_{10}\left[\frac{\left\|\mathbf{h} - \widehat{\mathbf{h}}(n)\right\|_2^2}{\|\mathbf{h}\|_2^2}\right] \tag{24}$$

as the normalized misalignment, Fig. 2(b) shows how the steady-state normalized misalignment degrades with increasing $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$. Hence, we note that as $\gamma \to 0$, $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}] \to 1$, a good misalignment performance is expected. As $\gamma \to 1$, $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}] \to \infty$ such that final normalized misalignment performance degrades significantly, as expected.

TABLE I
TWO-CHANNEL FREQUENCY-DOMAIN ADAPTIVE ALGORITHM [3]

μ'	=	$\mu(1-\lambda),\ 0 < \mu \le 2,\ 0 \ll \lambda < 1$
$\widehat{\mathbf{S}}_{jk}(m)$	=	$\lambda\widehat{\mathbf{S}}_{jk}(m-1) + (1-\lambda)\mathbf{D}_j^*(m)\mathbf{D}_k(m),\ j,k = 1,2$
$\lvert\widehat{\Gamma}(m)\rvert^2$	=	$\left[\widehat{\mathbf{S}}_{11}(m)\widehat{\mathbf{S}}_{22}(m)\right]^{-1}\widehat{\mathbf{S}}_{21}(m)\widehat{\mathbf{S}}_{12}(m)$
$\widehat{\mathbf{S}}_j(m)$	=	$\widehat{\mathbf{S}}_{jj}(m)\left[\mathbf{I}_{2L\times 2L} - \lvert\widehat{\Gamma}(m)\rvert^2\right],\ j = 1,2$
$\mathbf{K}_1(m)$	=	$\widehat{\mathbf{S}}_1^{-1}(m)[\mathbf{D}_1^*(m) - \widehat{\mathbf{S}}_{12}(m)\widehat{\mathbf{S}}_{22}^{-1}(m)\mathbf{D}_2^*(m)]$
$\mathbf{K}_2(m)$	=	$\widehat{\mathbf{S}}_2^{-1}(m)[\mathbf{D}_2^*(m) - \widehat{\mathbf{S}}_{21}(m)\widehat{\mathbf{S}}_{11}^{-1}(m)\mathbf{D}_1^*(m)]$
$\underline{\mathbf{e}}(m)$	=	$\underline{\mathbf{y}}(m) - \mathbf{G}^{01}_{2L\times 2L}\left[\mathbf{D}_1(m)\underline{\widehat{\mathbf{h}}}_1(m-1) + \mathbf{D}_2(m)\underline{\widehat{\mathbf{h}}}_2(m-1)\right]$
$\underline{\widehat{\mathbf{h}}}_j(m)$	=	$\underline{\widehat{\mathbf{h}}}_j(m-1) + \mu'\mathbf{K}_j(m)\underline{\mathbf{e}}(m),\ j = 1,2$

III. MISALIGNMENT OF SAEC ALGORITHM

We now examine how the formulation above can be applied to a frequency-domain adaptive algorithm to estimate its misalignment, hence verifying the relationship given in (23). It can be shown [3] that the steady-state normalized misalignment after convergence is approximated by

$$\eta(n) \approx 10\log_{10}\left[\frac{(1-\lambda)}{2}\frac{\sigma_w^2}{\|\mathbf{h}\|_2^2}\mathrm{tr}\{\mathbf{R}^{-1}\}\right] \tag{25}$$

such that for this two-channel case, we can express using (23)

$$\eta(n) \approx 10\log_{10}\left[\frac{(1-\lambda)4L}{2}\frac{\sigma_w^2}{\|\mathbf{h}\|_2^2\sigma_x^2}\chi^2_{\mathrm{E}}\left[\mathbf{R}^{\frac{1}{2}}\right]\right] \tag{26}$$

where σ_w^2 and $\sigma_x^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2$ are the noise and input signal variance, respectively. Hence, we note that the misalignment is a function of the forgetting factor $0 \ll \lambda < 1$, signal-to-noise ratio (SNR), and the condition number $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$.

The two-channel frequency-domain adaptive algorithm [3] has been shown to achieve good convergence performance for SAEC. For the jth channel, we first define the following quantities at each mth frame: $\mathbf{x}_j(m) = [x_j(mL-L), \ldots, x_j(mL+L-1)]^T$, $\mathbf{D}_j(m) = \mathrm{diag}\{\mathbf{F}_{2L\times 2L}\mathbf{x}_j(m)\}$, $\mathbf{W}^{01}_{2L\times 2L} = \begin{bmatrix}\mathbf{0}_{L\times L} & \mathbf{0}_{L\times L}\\ \mathbf{0}_{L\times L} & \mathbf{I}_{L\times L}\end{bmatrix}$, $\mathbf{G}^{01}_{2L\times 2L} = \mathbf{F}_{2L\times 2L}\mathbf{W}^{01}_{2L\times 2L}\mathbf{F}^{-1}_{2L\times 2L}$, $\underline{\mathbf{e}}(m) = \mathbf{F}_{2L\times 2L}\left[\begin{smallmatrix}\mathbf{0}_{L\times 1}\\ \mathbf{e}(m)\end{smallmatrix}\right]$, $\underline{\widehat{\mathbf{h}}}_j(m) = \mathbf{F}_{2L\times 2L}\left[\begin{smallmatrix}\widehat{\mathbf{h}}_j(m)\\ \mathbf{0}_{L\times 1}\end{smallmatrix}\right]$. The two-channel frequency-domain SAEC algorithm [3] is then given in Table I.

We note that for practicality, $\eta(n)$ can be computed using (25). However, in order to verify the validity of (23), we first compute $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$ using (23) such that elements $\gamma(l)$, $l = 0, \ldots, L-1$ are estimated using the interchannel coherence estimate $|\widehat{\Gamma}(m)|^2$ given in Table I. The theoretical steady-state normalized misalignment is then computed by employing (26). It is also evident from (23) and (26) that high interchannel coherence will degrade the conditioning of $\mathbf{R}$ and reduce the performance of the adaptive algorithm in terms of its final misalignment, as shown in Fig. 2(a) and (b). We also note that this formulation is more general than [13] since we have not assumed that $\mathbf{S}_{jj}$ is constant across frequency. This is feasible, especially for speech signals where the spectra is not constant across frequency, as will be shown in Section IV.

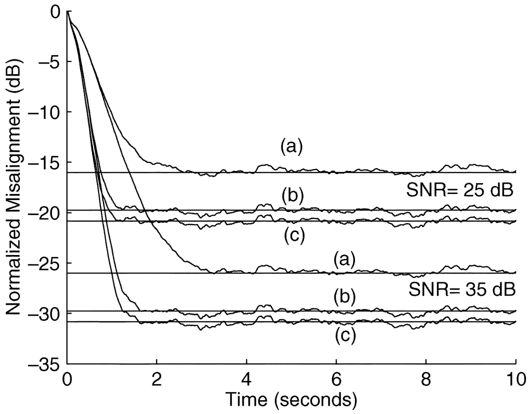

Fig. 3. Normalized misalignment for WGN input with mean interchannel coherences of (a) 0.85, (b) 0.60, and (c) 0.53.

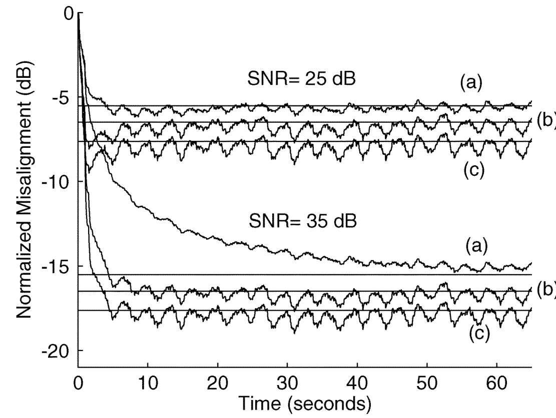

Fig. 4. Normalized misalignment for speech input with mean interchannel coherences of (a) 0.85, (b) 0.60, and (c) 0.53.

IV. Simulation Results

In our simulations, the lengths of both the adaptive filters are $L = 1024$ with $\lambda = [1 - 1/(3L)]^L$ and $\mu = 2$. The stereophonic impulse responses of both the transmission and receiving rooms are recorded at 16 kHz sampling rate and are of length 4096. To neglect any additional misalignment effects due to undermodeling, the impulse responses of the receiving room are truncated to 1024. Uncorrelated WGN $w(n)$ is added to achieve SNRs, as depicted in each experimental plot.

Fig. 3 shows the normalized misalignment plots for a WGN source input sequence such that $\mathbf{x}_1(n)$ and $\mathbf{x}_2(n)$ are generated by convolving this source with the impulse responses of the transmission room. We varied the interchannel coherence using a nonlinearity control factor β [1] such that

$$\mathbf{x}'_1(n) = \mathbf{x}_1(n) + 0.5\beta\left[\mathbf{x}_1(n) + |\mathbf{x}_1(n)|\right]$$
$$\mathbf{x}'_2(n) = \mathbf{x}_2(n) + 0.5\beta\left[\mathbf{x}_2(n) - |\mathbf{x}_2(n)|\right].$$

Theoretical steady-state normalized misalignments, shown as straight horizontal lines, are computed and averaged across block iterations using input signals $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$. Although we have chosen to use the nonlinearity control to vary the interchannel coherence, our analysis does not make any assumptions about the methods of achieving this variation. In order to verify (23), we computed the normalized misalignment using (23) and (25). Due to the variation of β, the measured mean interchannel coherences across frequency between $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$ are (a) 0.85, (b) 0.60, and (c) 0.53. We observe that as $\mathbf{x}'_1(n)$ and $\mathbf{x}'_2(n)$ become more uncorrelated, the final normalized misalignment reduces gracefully, as expected. We also note that the theoretical normalized steady-state misalignments computed using (23) are consistent with the experimental results, hence verifying the relationship between interchannel coherence and condition number of $\mathbf{R}$.

Fig. 4 shows normalized misalignment plots using speech input sequence from a male speaker. As before, the variation of interchannel coherence was controlled using β such that the measured mean interchannel coherences across frequency are the same as before. The mean theoretical normalized steady-state misalignments across time iterations are plotted as straight horizontal lines. We note that the normalized misalignment degrades with increasing interchannel coherence, as expected, and our theoretical normalized steady-state misalignment computed using $\chi^2_{\mathrm{E}}[\mathbf{R}^{1/2}]$ is consistent with the experimental results, hence verifying (23).

V. Conclusion

We derived the mathematical relationship between the condition number and interchannel coherence. This is achieved first by decomposing the covariance matrix in terms of the spectra of the input signal and then relating this spectra to the E-norm condition number. We have further shown how this relationship can be applied to estimate the steady-state normalized misalignment of a frequency-domain SAEC algorithm. Simulation results for both WGN and speech input have shown the validity of our theoretical analysis.

References

[1] 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," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 156–165, Mar. 1998.

[2] P. Eneroth, S. L. Gay, T. Gänsler, and J. Benesty, "A real-time implementation of a stereophonic acoustic echo canceller," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 513–523, Jul. 2001.

[3] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. New York: Springer-Verlag, 2001.

[4] A. Gilloire and V. Turbin, "Using auditory properties to improve the behavior of stereophonic acoustic echo cancellers," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 6, 1998, pp. 3681–3684.

[5] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 6, May 1998, pp. 3689–3692.

[6] A. W. H. Khong and P. A. Naylor, "Selective-tap adaptive algorithms in the solution of the nonuniqueness problem for stereophonic acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 269–272, Apr. 2005.

[7] A. Sugiyama, A. Hirano, and K. Nakayama, "Acoustic echo cancellation for conference systems," in *Proc. Eur. Signal Processing Conf.*, 2004, pp. 17–20.

[8] J. Benesty and T. Gänsler, "A recursive estimation of the condition number in the RLS algorithm," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 4, 2005, pp. 25–28.

[9] H. Buchner, J. Benesty, and W. Kellermann, "Generalized multichannel frequency-domain adaptive filtering: Efficient realization and application to hands-free speech communication," *Signal Process.*, vol. 85, pp. 549–570, Mar. 2005.

[10] S. Haykin, *Adaptive Filter Theory*, 4th ed, ser. Information and System Science. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[11] R. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inf. Theory*, vol. 18, no. 6, pp. 725–730, Nov. 1972.

[12] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[13] T. Gänsler and J. Benesty, "New insights into the stereophonic acoustic echo cancellation problem and an adaptive nonlinearity solution," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 257–267, Sep. 2002.