On the Use of Channel Shortening in Multichannel Acoustic System Equalization

Wancheng Zhang, Emanuël A. P. Habets, and Patrick A. Naylor

{wancheng.zhang07, e.habets, p.naylor}@imperial.ac.uk

Department of Electrical and Electronic Engineering, Imperial College London
London, SW7 2AZ, United Kingdom

***Abstract*—This work investigates the use of channel shortening (CS) in equalization of acoustic systems. Channel shortening techniques have been extensively developed in the context of digital communications and applied to acoustic systems. However, at this point of time, the use of CS for acoustic system equalization has not been well established. A mathematical link between the MINT and the traditional CS is derived. In multichannel scenarios, CS can provide multiple solutions but not all of them are useful in the sense of perceptual quality. Therefore, a criterion for spelling out a perceptually advantageous equalization system from the multiple solutions to CS is provided. Relaxed multichannel least-squares (RMCLS) method is presented and simulations confirm that the RMCLS outperforms CS in the presence of channel estimation errors.**

I. Introduction

System equalization can be potentially used to remove the adverse effects of reverberation cause by a convolutive acoustic channel between a talker and the receiving microphone. This work considers the case when the acoustic channel impulse response has already been estimated and the task at hand is to employ this estimate in order to design an equalizer that will improved the perceived quality of the speech signal. A typical application of this technique is for speech enhancement of hands-free speech captured in reverberant rooms. Channel shortening (CS) techniques have been extensively developed in the context of digital communications to mitigate the inter-symbol and inter-carrier interference. Both closed-form [1] and adaptive [2]–[4] methods have been well studied. These techniques have been extended for multiple-input/multiple-output (MIMO) systems in [5], [6]. A common frame work for CS can be found in [7]. Channel shortening has been used for acoustic system equalization in [8] and [9]. In previous work, the main issue of concern is that the channel impulse response is shortened; the coefficients of the impulse response after shortening are of no concern. In acoustic system equalization however, the coefficients of the shortened impulse response are important in that they can lead to different perceptual sound quality. In their latest work presented in [10], the authors considered the psychoacoustic property of masking effects in shortening single-channel impulse response.

In this work, we discuss the use of channel shortening with respect to perceptual sound quality for multichannel systems. Before Section V, true impulse responses without estimation error are assumed to be known. In Section V, the robustness of CS to channel estimation errors is considered.

Consider an M-channel acoustic system $\mathbf{h} = [\mathbf{h}_1^T \ \cdots \ \mathbf{h}_M^T]^T$. The acoustic channel between the source and the mth microphone is characterized by its impulse response $\mathbf{h}_m = [h_m(0) \ h_m(1) \ \ldots \ h_m(L-1)]^T$, $m = 1, \ldots, M$, where $\{\cdot\}^T$ denotes the transpose operation. In our work we assume that the acoustic channels are time-invariant. In a unified form [7], which is adopted in [9], CS aims to use a multichannel equalization system $\mathbf{g} = [\mathbf{g}_1^T \ \mathbf{g}_2^T \ \cdots \ \mathbf{g}_M^T]^T$, where $\mathbf{g}_m = [g_m(0) \ g_m(1) \ \ldots \ g_m(L_i-1)]^T$ is the mth component of length L_i, to maximize a generalized Rayleigh quotient

$$\mathbf{g} = \arg\max_{\mathbf{g}} \frac{\mathbf{g}^T\mathbf{B}\mathbf{g}}{\mathbf{g}^T\mathbf{A}\mathbf{g}}, \tag{1}$$

where

$$\begin{aligned}\mathbf{B} &= \mathbf{H}^T \mathrm{diag}\{\mathbf{w}_\mathrm{d}\}^T \mathrm{diag}\{\mathbf{w}_\mathrm{d}\}\mathbf{H} \\ \mathbf{A} &= \mathbf{H}^T \mathrm{diag}\{\mathbf{w}_\mathrm{u}\}^T \mathrm{diag}\{\mathbf{w}_\mathrm{u}\}\mathbf{H}\end{aligned}$$

with $\mathbf{H} = [\mathbf{H}_1 \ \cdots \ \mathbf{H}_M]$ and $\mathbf{H}_m$ the $(L + L_i - 1) \times L_i$ convolution matrix of $\mathbf{h}_m$, and

$$\begin{aligned}\mathbf{w}_\mathrm{d} &= [\underbrace{0 \ \cdots \ 0}_{\tau} \ \underbrace{1 \ \cdots \ 1}_{L_w} \ 0 \ \cdots \ 0]^T_{[L+L_i-1]} \\ \mathbf{w}_\mathrm{u} &= \mathbf{1}_{[(L+L_i-1)\times 1]} - \mathbf{w}_\mathrm{d},\end{aligned}$$

where τ is an optional integer delay, and L_w defines the region that is maximized and will be referred to as the *window area*. In contrast, the region outside the window area will be referred to as *wall area*. For the sake of discussion, here we define the equalized impulse response (EIR), which is the impulse response from the source to the output of the equalization system,

$$b(i) = \sum_{m=1}^{M} h_m(i) * g_m(i), \tag{2}$$

where $*$ denotes linear convolution.

For single-channel system shortening, since the target shortening length L_w is always desired to be smaller than the channel length L, $\mathbf{A}$ is of full rank. For multichannel shortening and for specific design parameters settings, $\mathbf{A}$ can be rank deficient. For a rank deficient $\mathbf{A}$, (1) has multiple solutions. Any of these solutions leads to an EIR containing only zeros in the wall area, but different solutions lead to EIRs of different window area coefficients. In digital communications, the main issue of concern is that the Rayleigh quotient in (1) is maximized; the coefficients in the window area after shortening are not important. In acoustic system equalization however, although all solutions maximize the Rayleigh quotient, the resulting EIRs are different from a perceptual point of view. Some of them result in improved perceptual sound quality after equalization compared to before, but some do not.

In this paper, firstly, a mathematical link between the CS and the well-known multiple-input/output inverse theorem (MINT) [11] is derived. Next, a criterion for selecting perceptually advantageous equalization system from the multiple solutions to CS is provided. Then, an alternative approach to CS, which we call relaxed multichannel least-squares (RMCLS), is proposed. After this, since in practice the true channel impulse responses are not available and what we have are their estimates, the robustness of MINT, CS, and RMCLS to channel estimation errors is investigated.

II. A LINK BETWEEN MINT AND CHANNEL SHORTENING

As shown by MINT, solution(s) to the system of equations

$$\mathbf{H}\mathbf{g} = \mathbf{d}, \tag{3}$$

where

$$\mathbf{d} = [\underbrace{0\ \cdots\ 0}_{\tau}\ 1\ 0\ \ldots\ 0]^T_{[L+L_i-1]}, \tag{4}$$

exist when the following two conditions are both satisfied:

C-1 $H_m(z^{-1})$, the z-transforms of the multichannel impulse responses $\mathbf{h}_m$ do not share any common zeros [11].

C-2 The length of $\mathbf{g}_m$ $L_i \geq L_c$ [12][1], where

$$L_c = \left\lceil \frac{L-1}{M-1} \right\rceil \tag{5}$$

where $\lceil \kappa \rceil$ denotes the smallest integer larger than or equal to κ.

When both C-1 and C-2 are satisfied, $\mathbf{A}$ is rank deficient. Therefore, $\mathbf{g}$ that satisfies

$$\begin{cases} \mathbf{g}^T\mathbf{A}\mathbf{g} &= 0 \\ \mathbf{g}^T\mathbf{B}\mathbf{g} &\neq 0 \end{cases} \tag{6}$$

maximizes the Rayleigh quotient in (1). Equivalently, since $\mathbf{F} \doteq \mathbf{H}^T\mathbf{H} = \mathbf{B} + \mathbf{A}$, $\mathbf{g}$ that satisfies

$$\begin{cases} \mathbf{g}^T\mathbf{A}\mathbf{g} &= 0 \\ \mathbf{g}^T\mathbf{F}\mathbf{g} &\neq 0 \end{cases} \tag{7}$$

is a solution to (1).

The matrix $\mathbf{H}$ is full row-rank [12], and the rank of $\mathbf{F}$ is $(L + L_i - 1)$. Since

$$\dim(\mathrm{null}(\mathbf{F})) = ML_i - (L + L_i - 1) \doteq L_F, \tag{8}$$

where $\dim(\cdot)$ denotes the dimension and $\mathrm{null}(\cdot)$ denotes the null space, we assume vectors $\mathbf{p}^1,\ \mathbf{p}^2,\ \ldots,\ \mathbf{p}^{L_F}$ to be a basis of $\mathrm{null}(\mathbf{F})$. Since

$$\dim(\mathrm{null}(\mathbf{A})) = ML_i - (L + L_i - 1 - L_w) \tag{9}$$

and $\mathrm{null}(\mathbf{F}) \subset \mathrm{null}(\mathbf{A})$, we can assume $\mathbf{g}_{\mathrm{CS}}^1$, $\mathbf{g}_{\mathrm{CS}}^2$, …, $\mathbf{g}_{\mathrm{CS}}^{L_w}$, $\mathbf{p}^1$, $\mathbf{p}^2$, …, $\mathbf{p}^{L_F}$ to be a basis of $\mathrm{null}(\mathbf{A})$. Any solution to (1), which is in the space $\mathrm{null}(\mathbf{A})\backslash\mathrm{null}(\mathbf{F})$, where $\backslash$ denotes exclusion operator, can be expressed as

$$\mathbf{g} = [\mathbf{g}_{\mathrm{CS}}^1\ \mathbf{g}_{\mathrm{CS}}^2\ \cdots\ \mathbf{g}_{\mathrm{CS}}^{L_w}\ \mathbf{p}^1\ \mathbf{p}^2\ \ldots\ \mathbf{p}^{L_F}] \begin{bmatrix} \mathbf{t}_{[L_w\times 1]} \\ \mathbf{r}_{[L_F\times 1]} \end{bmatrix}, \tag{10}$$

with $\mathbf{t}_{[L_w\times 1]} \neq \mathbf{0}$.

In [7], the $\mathbf{g}$ maximizing the quotient in (1) is found by solving the the generalized eigenvalue problem

$$\mathbf{B}\mathbf{g} = \lambda\mathbf{A}\mathbf{g}. \tag{11}$$

The solution to (1) is obtained by computing the eigenvector relating to the largest eigenvalue. For rank deficient $\mathbf{A}$, (11) can be solved using the QZ algorithm [14]. In this work, the MATLAB function $\mathrm{eig}(B, A, \text{‘qz’})$ which employs the QZ algorithm is used to solve (11). L_w vectors relating to $\lambda = \infty$ can be obtained. Without loss of generality, we assume the L_w vectors obtained from solving the generalized eigenvalue problem are $\mathbf{g}_{\mathrm{CS}}^1$, $\mathbf{g}_{\mathrm{CS}}^2$, …, $\mathbf{g}_{\mathrm{CS}}^{L_w}$. Since the equalization system(s) obtained from MINT satisfy (7) as well, there must be one MINT solution which can be expressed as a linear combination of these L_w vectors. We have

$$\mathbf{H}[\mathbf{g}_{\mathrm{CS}}^1\ \mathbf{g}_{\mathrm{CS}}^2\ \cdots\ \mathbf{g}_{\mathrm{CS}}^{L_w}] = \begin{bmatrix} \mathbf{0}_{[\tau\times L_w]} \\ \mathbf{D}_{[L_w\times L_w]} \\ \mathbf{0}_{[(L+L_i-1-\tau-L_w)\times L_w]} \end{bmatrix}, \tag{12}$$

where $\mathbf{D}$ is a full-rank matrix. Letting

$$\mathbf{q} = \mathbf{D}^{-1}\mathbf{d}', \tag{13}$$

where $\mathbf{d}' = [1\ 0\ \ldots\ 0]^T_{[L_w]}$, we have

$$\mathbf{H}[\mathbf{g}_{\mathrm{CS}}^1\ \mathbf{g}_{\mathrm{CS}}^2\ \cdots\ \mathbf{g}_{\mathrm{CS}}^{L_w}]\mathbf{q} = \mathbf{d}. \tag{14}$$

Therefore, $[\mathbf{g}_{\mathrm{CS}}^1\ \mathbf{g}_{\mathrm{CS}}^2\ \cdots\ \mathbf{g}_{\mathrm{CS}}^{L_w}]\mathbf{q}$, which is a linear combination of the L_w vectors obtained from channel shortening, is an MINT solution.

III. A CRITERION FOR SELECTING A PERCEPTUALLY ADVANTAGEOUS EQUALIZATION SYSTEM

The MATLAB function $\mathrm{eig}(B, A, \text{‘qz’})$ provides L_w independent vectors corresponding to $\lambda = \infty$. Using these vectors as equalization systems, some of them result in EIRs which lead to improved perceptual speech quality after equalization compared to the original received signal, but some do not.

Figure 1(a) and (b) shows two EIRs resulting from two equalization systems of the L_w solutions obtained by solving the generalized eigenvalue problem, for which a 2-channel system from MARDY database [15] with $L = 2000$ is used and design parameters are set to $L_i = L_c$, $\tau = 0$, and $L_w = 400$ (corresponding to 50 ms for sampling frequency $f_s = 8$ kHz, which is a typical transition time between early reflections and late reflections for room impulse responses

[1] It should be noted that solution(s) to (3) always exist when $L_i \geq L - 1$ [13]. Unfortunately, this cannot be guaranteed when $L_i \geq L_c$. However, it has been proved in [12] that solution(s) exist for almost all cases.

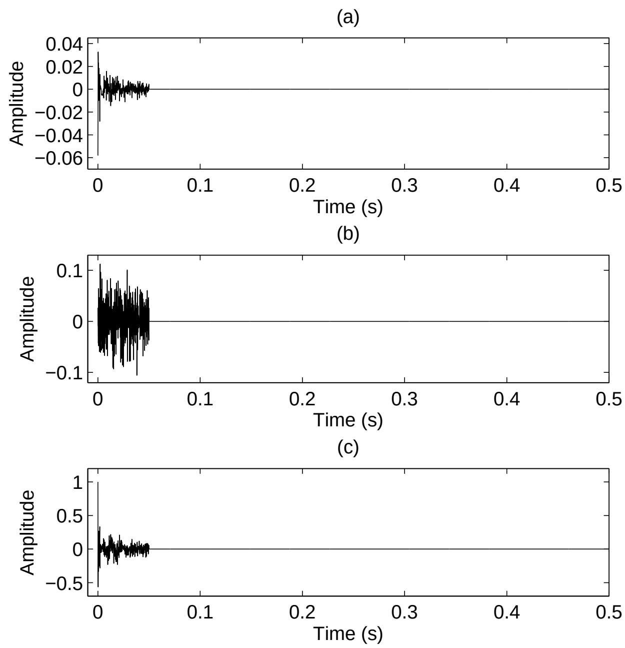

Fig. 1. (a) An EIR obtained from CS resulting in perceptually improved speech, (b) an EIR obtained from CS resulting in perceptually degraded speech, and (c) EIR obtained from RMCLS. Note that all the EIRs equal 0 after 0.05 s.

(RIRs)). Perceptually, for an RIR, early reflections are not perceived as separate sound events but instead cause a spectral distortion called colouration, whereas late reflections often form a background ambience which is distinct from the foreground sound and may impair speech intelligibility [16]. Although these two EIRs in Fig. 1 only have early reflections, they may not be both satisfactory for perception. It can be seen that the window area shown in Fig. 1(a) has a decaying pattern which is similar to that of an RIR, whereas the one shown in Fig. 1(b) has a non-decaying pattern. We conducted informal listening tests to investigate the sound quality of speech relating to the different EIRs. In the tests, speech segments were listened using a Sennheiser HD 650 headphone by 8 subjects. The following questions were investigated in the tests: is the speech warm or not warm, thin or not thin, perceptually close to or not close to the anechoic speech? The obtained results indicated that speech relating to the EIR in Fig. 1(b) is perceived as thin and harsh, whereas speech relating to the EIR in Fig. 1(a) is perceived as warm, closer to the anechoic speech, and is perceptually preferred. The speech resulting from some other solutions obtained from the generalized eigenvalue decomposition sounds similar to that relating to the EIR in Fig. 1(b) and the solution leading to the EIR in Fig. 1(a) is preferred over these solutions. An exhaustive comparison of all L_w solutions is not necessary when a criterion is available with which the solution leading to the EIR in Fig. 1(a) can be picked out. We found that the solution leading to the EIR in Fig. 1(a) has the following characteristic: among the EIRs resulted from all L_w solutions, the EIR resulted from it has minimum ℓ_2-norm.

Experiments with 30 different acoustic systems show that an equalization system leading to an EIR similar (in terms of leading to similar sounding equalized speech) to the one in Fig. 1(a) always exists for different systems, and it always retains the above ℓ_2-norm characteristic. Therefore, we propose the following criterion for selecting a solution that will result in improved perceptual quality of the speech: among the multiple solutions obtained by solving the generalized eigenvalue problem (11), we choose the one which results in the minimum ℓ_2-norm EIR. In the remainder of this work, when CS equalization system is mentioned, it always refers to the one selected with the above criterion.

IV. Relaxed Multichannel Least-Squares Method for Channel Shortening

In this Section, an alternative approach to CS is presented. We propose to achieve CS by minimizing the following cost function

$$J = \|\mathbf{W}(\mathbf{H}\mathbf{g} - \mathbf{d})\|_2^2, \tag{15}$$

where $\mathbf{W} = \mathrm{diag}\{\mathbf{w}\}$ with

$$\mathbf{w} = [\underbrace{1 \ \ldots \ 1}_{\tau} \ \underbrace{1 \ 0 \ \ldots \ 0}_{L_w} \ 1 \ \ldots \ 1]^T_{(L+L_i-1)\times 1}. \tag{16}$$

The first weight in the window area in $\mathbf{w}$ is set to 1 so that the trivial solution can be avoided. Since by using the weighting function $\mathbf{w}$ the minimization in the window area is relaxed, it is called relaxed multichannel least-squares (RMCLS). The solution obtained from

$$\mathbf{g} = (\mathbf{W}\mathbf{H})^{+}\mathbf{W}\mathbf{d} \tag{17}$$

is used as the equalization system, where $\{\cdot\}^+$ denotes Moore-Penrose pseudo-inverse [17]. This solution is the minimum ℓ_2-norm solution among all possible solutions minimizing (15) or maximizing the quotient in (1), and is also in the solution space of (1).

The EIR given by the RMCLS is shown in Fig. 1(c). It can be seen that in the window area, it also shows a decaying pattern. In informal listening tests, the sound of speech resulting from RMCLS is thinner than the selected CS equalization system, but is warmer and preferred to the other CS solutions.

V. Equalization in the Presence of Channel Estimation Error

In practice, we never know the true RIRs and equalization systems can only be computed based on the estimates of the RIRs. Since the acoustic channel estimates always include errors, the equalization system computed from the estimates might not be able to equalize the true system. In this Section, we show the performance of MINT, CS, and RMCLS in the presence of the error by a simulation example. The 2-channel system used above is employed as the true system and is estimated using the system identification method described in

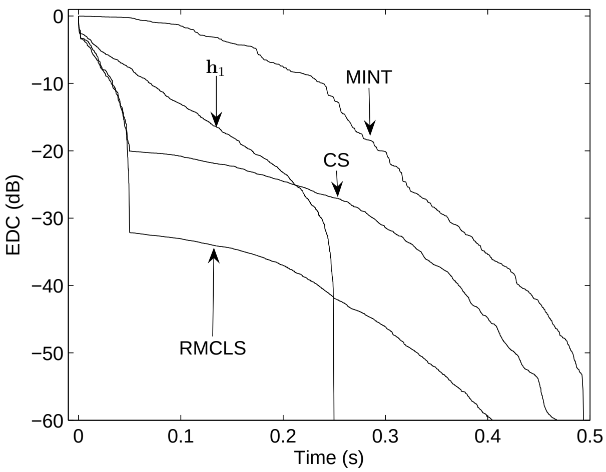

Fig. 2. The EDC of $\mathbf{h}_1$, and the EDCs of the EIRs obtained from MINT, CS, and RMCLS in the presence of channel estimation error. Note that in the absences of estimation errors, the EDC for MINT equals $-\infty$ dB after the first tap, and the EDCs for CS and RMCLS equal $-\infty$ dB after 0.05 s.

[18, pp. 90-94], which is a supervised system identification method. In this example, the normalized level of the error, $(\|\mathbf{h}-\hat{\mathbf{h}}\|_2^2)/\|\mathbf{h}\|_2^2$, is -33 dB, where $\hat{\mathbf{h}}$ is the estimate of $\mathbf{h}$. The energy decay curves (EDCs) [16] of the EIRs are shown in Fig. 2. It can be seen that in the presence of estimation error, the MINT totally fails to equalize the acoustic system. For CS equalization system, the part of the EDC before 0.2 s is below the EDC of $\mathbf{h}_1$. However, the decay rate of the EDC after 0.05 s is smaller than that of $\mathbf{h}_1$, and a deleterious tail in the EIR is introduced. On the other hand, the RMCLS equalization system is superior in robustness to estimation error. It can be seen that the RMCLS is more robust than the CS. The EDCs for RMCLS shows more than 15 dB reduction compared with $\mathbf{h}_1$ at any time after 0.05 s and the artificial tail is well below a level that can be perceived by the listener.

VI. Conclusion

In this work, the use of channel shortening technique in equalization of acoustic systems is investigated. A mathematical link between MINT and CS is derived. Multiple solutions to CS can be obtained and one MINT solution can be expressed as a linear combination of the CS solutions. A criterion for selecting a perceptually advantageous equalization system from the multiple solutions to CS is provided. The results of our informal listening tests showed that equalization using the solution corresponding to the EIR with minimum ℓ_2-norm is perceptually preferred. Extended listening tests are required to reach scientifically significant conclusions. RMCLS algorithm is presented, and the performance of CS is compared with RMCLS. Simulations show that RMCLS outperforms CS in robustness to channel estimation errors.

Acknowledgment

The authors acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 226007 SCENIC.

References

[1] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 1662–1672, 1996.

[2] M. Nafie and A. Gatherer, "Time-domain equalizer training for ADSL," in *IEEE Int. Conf. Commun.*, 1997.

[3] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson Jr., "A blind, adaptive TEQ for multicarrier systems," *IEEE Signal Processing Lett.*, vol. 9, pp. 341–343, 2002.

[4] R. Nawaz and J. A. Chambers, "Blind adaptive channel shortening by single lag autocorrelation minimisation," *Electronics letters*, vol. 40, no. 25, 2004.

[5] N. Al-Dhahir, "FIR channel shortening equalizers for MIMO ISI channels," *IEEE Trans. Commun.*, vol. 49, pp. 213–218, 2001.

[6] R. K. Martin, J. M. Walsh, and C. R. Johnson Jr., "Low-complexity MIMO blind, adaptive channel shortening," *IEEE Trans. Signal Processing*, vol. 53, pp. 1324–1334, 2005.

[7] R. K. Martin, K. Vanbleu, M. Ding, G. Ysebaert, M. Milosevic, B. L. Evans, M. Moonen, and C. R. Johnson Jr., "Unification and evaluation of equalization structures and design algorithms for discrete multitone modulation systems," *IEEE Trans. Signal Process.*, vol. 53, pp. 3880–3894, 2005.

[8] M. Kallinger and A. Mertins, "Impulse response shortening for acoustic listening room compensation," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2005.

[9] ——, "Multi-channel room impulse response shaping - a study," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2006.

[10] A. Mertins, T. Mei, and M. Kallinger, "Room impulse response shortening/reshaping with infinity- p-norm optimization," *IEEE Trans. Audio, Speech, Language Processing*, vol. 18, pp. 249–259, 2010.

[11] M. Miyoshi and K. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 145–152, 1988.

[12] G. Harikumar and Y. Bresler, "FIR perfect signal reconstruction from multiple convolutions: minimum deconvolver orders," *IEEE Trans. Signal Process.*, vol. 46, pp. 215–218, 1998.

[13] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, pp. 516–525, 1995.

[14] C. B. Moler and G. W. Stewart, "An algorithm for generalized matrix eigenvalue problems," *SIAM J. Number. Anal.*, vol. 10, no. 2, pp. 241–256, 1973.

[15] J. Y. C. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2006.

[16] H. Kuttruff, *Room acoustics*, 4th ed. Taylor & Frances, 2000.

[17] R. Rado, "Note on generalized inverse of matrices," *Proc. Cambridge Philos. Soc.*, vol. 52, pp. 600–601, 1956.

[18] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993, vol. 1.