

COMPUTATIONALLY EFFICIENT EQUALIZATION OF ROOM IMPULSE RESPONSES ROBUST TO SYSTEM ESTIMATION ERRORS

Wancheng Zhang, Nikolay D. Gaubitch and Patrick A. Naylor

Imperial College London, Department of Electrical and Electronic Engineering

ABSTRACT

Equalization techniques for room impulse responses (RIRs) are important in acoustic signal processing applications such as speech dereverberation. In practice, only approximate estimates of the RIRs are available and the inverse filters designed from these estimates may cause significant distortion in the equalized signal. A second issue is that existing equalizer design algorithms are computationally expensive. We here propose regularized subband equalizer design algorithm. Both the computational complexity and the robustness of the equalizer design to system estimation errors are improved. An analysis of the computational complexity and simulation examples are provided to support our study.

Index Terms— acoustic signal processing, deconvolution, equalizers, inverse problems

1. INTRODUCTION

Equalization of room impulse responses (RIRs) is an important research topic with several applications in acoustic signal processing, such as hands free telephony and automatic speech recognition [1]. We focus here on the case of a single source (talker) and $M \geq 2$ microphones. Such multichannel equalization is often preceded by a process of identifying the acoustic channels between the source and the receivers.

Consider the L -tap room impulse response of the acoustic path between a source and the m th microphone, $\mathbf{h}_m = [h_{m,0}, h_{m,1}, \dots, h_{m,L-1}]^T$. Equalization can be achieved by an inverse system, $\mathbf{g}_m = [g_{m,0}, g_{m,1}, \dots, g_{m,L_i-1}]^T$, satisfying

$$\mathbf{H}_m \mathbf{g}_m = \mathbf{d}, \quad (1)$$

where \mathbf{H}_m is an $(L + L_i - 1) \times L_i$ convolution matrix,

$$\mathbf{d} = [\underbrace{0, \dots, 0}_\tau, 1, 0, \dots, 0]^T \quad (2)$$

is an $(L + L_i - 1) \times 1$ vector, τ is a modeling delay and L_i is the design length of the inverse filter. However, because of the non-minimum phase nature of the RIRs [2], stable causal solutions for \mathbf{g}_m can not normally be obtained. Exact equalization is indeed possible when the MINT [3] algorithm is

used with multiple microphones, but there are several practical problems with RIR equalization that remain unsolved [4, 5]. Room impulse response estimation errors can be large due to, for example, measurement noise. Equalization filters designed from inaccurate estimates of RIRs can cause distortion in the equalized signal [4, 6]. Furthermore, existing equalization algorithms [3, 7, 8, 9, 10] are computationally complex, which also causes problems for practical implementations.

An oversampled subband multichannel least squares (LS) equalization algorithm is proposed in [6] and it is shown that equalization in subbands can substantially decrease computational complexity compared with the fullband counterpart. In [9], the authors use regularization to reduce the sensitivity of inverse filters to RIR fluctuations and observation noise. In this paper, we propose the use of regularization in the subband algorithm in order to improve the robustness of the design to RIR estimation errors.

The remainder of the paper is organized as follows. In Section 2, regularized LS equalization and subband approaches are reviewed. A computational complexity analysis is given in Section 3. In Section 4, a comparison of the robustness between two approaches is provided and conclusions are drawn in Section 5.

2. MULTICHANNEL LS EQUALIZATION

We now review the regularized multichannel equalization algorithm [9] and introduce it into the subband implementation [6].

2.1. Regularized Multichannel LS Equalization

A multichannel inverse filter $\mathbf{g} = [\mathbf{g}_1^T, \mathbf{g}_2^T, \dots, \mathbf{g}_M^T]^T$ can be obtained by solving the equation,

$$\mathbf{H} \mathbf{g} = \mathbf{d}, \quad (3)$$

where $\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_M]$ is the system matrix. The inverse filter can be obtained by

$$\mathbf{g} = \mathbf{H}^+ \mathbf{d}, \quad (4)$$

where $\mathbf{H}^+ = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ denotes the Moore-Penrose pseudo-inverse of matrix \mathbf{H} .

The effectiveness in increasing the robustness of the design of \mathbf{g} to RIR fluctuations and observation noise as a function of regularization parameter δ , filter length L_i , and modeling delay τ , is investigated in [9]. The regularized LS equalization filter can be calculated according to

$$\mathbf{g} = (\mathbf{H}^T \mathbf{H} + \delta \mathbf{I})^{-1} \mathbf{H}^T \mathbf{d}, \quad (5)$$

where \mathbf{I} is an identity matrix. Regularization reduces the accuracy of the inverse filter, but it can also reduce the sensitivity to RIR estimation errors in \mathbf{H} . For this equalization filter, a reference length of $L_i = \lceil \frac{L-1}{M-1} \rceil$ is used, which makes the matrix \mathbf{H} square when $\frac{L-1}{M-1}$ is an integer and thus an exact inversion can be obtained if the channels do not share any common zeros [3]. However, improved equalization can be expected by increasing the length of the equalization filter or choosing appropriate delay τ in (2) when estimation errors are present.

2.2. Subband Multichannel LS Equalization

For the case of subband multichannel LS equalization [6], the generalized discrete Fourier transform (GDFT) filterbank [11] is employed. The reverberant speech is first decomposed into K subbands with L_{pr} -tap analysis filters \mathbf{u}_k and decimated by a factor of N . Subband equalization is then performed using the approach already described in the fullband context. The equalized signal is finally reconstructed with synthesis filters.

In order to design the subband equalizer, the subband RIRs must be found from fullband estimates using, for example, complex subband decomposition [12]. The estimates of the k th subband, m th channel RIRs are found by solving the following optimization problem

$$\mathbf{h}'_{k,m} = \arg \min_{\mathbf{h}'_{k,m}} \|\mathbf{U}_{N,k} \mathbf{h}'_{k,m} - \mathbf{r}_{N,km}\|_2^2, \quad (6)$$

where $\mathbf{U}_{N,k}$ is the convolution matrix of the decimated analysis filters $\mathbf{u}_{N,k} = [u_{k,0}, u_{k,N}, \dots, u_{k,L_{pr}-1}]^T$ and $\mathbf{r}_{N,km} = [r_{k,m,0}, r_{k,m,N}, \dots, r_{k,m,N(L-1)}]^T$ is an $\lceil \frac{L+L_{pr}-1}{N} \rceil \times 1$ vector with $\mathbf{r}_{k,m} = \mathbf{h}_m * \mathbf{u}_k$. Then the channel filters are obtained in the least squares optimal sense according to

$$\mathbf{h}'_{k,m} = \mathbf{U}_{N,k}^+ \mathbf{r}_{N,km}, \quad (7)$$

with length of $L' = \lceil \frac{L+L_{pr}-1}{N} \rceil - \lceil \frac{L_{pr}}{N} \rceil + 1$.

By introducing regularization into the subband equalization, the equalization filters \mathbf{g}_k of k th subband RIRs \mathbf{h}'_k can be calculated according to

$$\mathbf{g}_k = (\mathbf{H}_k^T \mathbf{H}_k + \delta \mathbf{I})^{-1} \mathbf{H}_k^T \mathbf{d}', \quad (8)$$

where $\mathbf{H}_k = [\mathbf{H}_{k,1}, \mathbf{H}_{k,2}, \dots, \mathbf{H}_{k,M}]$ is the k th subband system matrix, $\mathbf{d}' = [0, \dots, 0, 1, 0, \dots, 0]^T$ is an $(L' + L'_i - 1) \times 1$ vector, $\mathbf{H}_{k,m}$ is the convolution matrix of $\mathbf{h}_{k,m}$ and L'_i is the length of the subband equalization filters with a reference value of $L'_i = \lceil \frac{L'-1}{M-1} \rceil$.

3. COMPUTATIONAL COMPLEXITY ANALYSIS

In this section we present a comparative analysis of the computations required for the fullband and subband equalization algorithms. The comparison is made in terms of floating point operations (flops) [13]. We consider the general optimization problem $\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$, which has an LS solution $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$, where \mathbf{A} is an arbitrary real valued $p \times q$ matrix and \mathbf{b} is a real valued $p \times 1$ vector. The number of flops required to solve this problem is given by [13]

$$pq^2 + \frac{q^3}{3}. \quad (9)$$

From the dimension of the fullband equalization filter calculation in (5), the number of flops required is

$$(ML_i)^2(L + L_i - 1) + \frac{(ML_i)^3}{3}. \quad (10)$$

The subband equalization filter design takes into consideration two separate calculations for each of the $\frac{K}{2}$ subbands: the cost of the subband decomposition in (7) and the equalization filter computation in (8). The data for these calculations is complex where, generally, one complex multiply requires four real multiplies and two real additions, and one complex addition requires two real additions. Thus, we multiply the expression in (9) by a factor of six. The total flops required for the subband equalization filter design can be expressed as

$$3K(ML'_i)^2(L' + L'_i - 1) + L'^2 \left(\frac{L_{pr}}{N} + L' - 1 \right) + \frac{(ML'_i)^3 + L'^3}{3}. \quad (11)$$

For the case of the reference equalization filter length used, the ratio of (10) and (11) is given by

$$\frac{N^3 LM^3 (M + 2)}{3KLM^3(M + 2) + (4L + 3L_{pr})(M - 1)}. \quad (12)$$

4. EXPERIMENTS

Simulations are used to demonstrate the subband regularization approach to the equalizer design robustness to estimation errors in the RIRs. Figure 1 shows the configuration of the source and $M = 2$ microphones used in the experiments, following the scenario used in [14]. Room impulse responses are generated using the image method [15]. The sampling frequency is 8 kHz and the reverberation time is 200 ms, which results in RIRs of 1600 taps.

The channel estimation error of the i th tap, m th channel is modeled as [16]

$$\xi_{m,i} = \epsilon_{m,i} \times h_{m,i}, \quad (13)$$

where $\epsilon_{m,i}$ is a zero mean Gaussian variable with variance set to desired system mismatch $E_m = 10 \log_{10}(\text{var}(\epsilon_{m,i}))$ dB, such that $E_m = 10 \log_{10} \frac{\|\xi_m\|_2^2}{\|\mathbf{h}_m\|_2^2}$ dB.

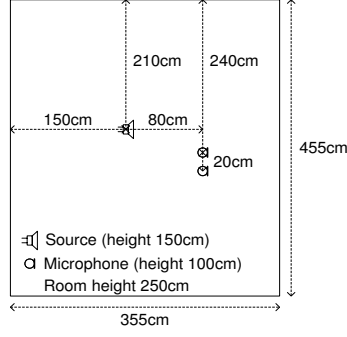


Fig. 1. Room dimensions and source-microphone configuration.

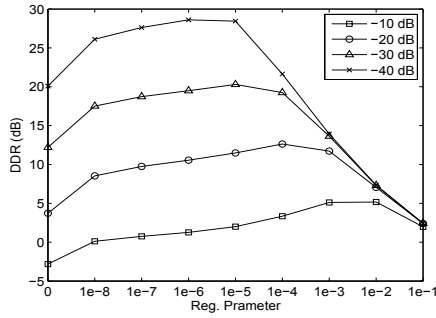


Fig. 2. Performance as a function of regularization parameter for system mismatch $E_m = -10, -20, -30$, and -40 dB.

Then, the equalization filter for the fullband and subband cases are calculated with (5), and (8) respectively using the RIR estimates $\hat{\mathbf{h}} = \mathbf{h} + \boldsymbol{\xi}$. Finally, we use this filter to equalize the room impulse responses.

The filterbank used in this paper uses $K = 32$ subbands and decimation factor, $N = 24$, as used in [6]. An $L_{pr} = 512$ -tap prototype filter was designed using the iterative least squares method [11].

For quantitative evaluation, we use the following metrics:

a. Direct to Reverberant Ratio (DDR)—The direct path impulse response is defined as \mathbf{h}_d representing propagation from the talker to a microphone without reflections, and the reverberant component \mathbf{h}_r as an impulse response representing all nondirect propagation paths. The DDR can be found from the corresponding impulse response coefficients [5] as

$$\text{DDR} = 10 \log_{10} \frac{\|\mathbf{h}_d\|_2^2}{\|\mathbf{h} - \mathbf{h}_d\|_2^2} \text{ dB} \quad (14)$$

b. Magnitude and Linear phase deviation—We define $D(f) = |D(f)|e^{j\theta(f)}$ as the F -point discrete Fourier transform of the equalized impulse response. Magnitude deviation is defined as the standard deviation of the equalized magni-

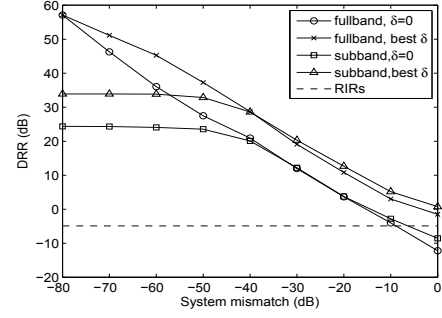


Fig. 3. Comparison in DDR

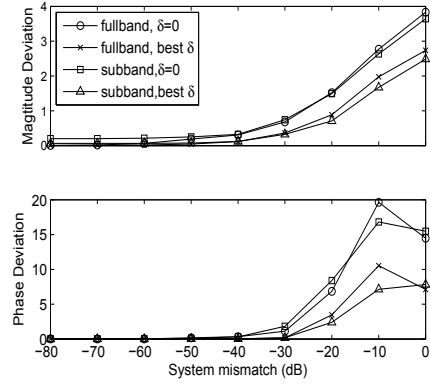


Fig. 4. Comparison in Magnitude and Linear phase deviation

tude response [4]

$$\sigma = \sqrt{\frac{1}{F} \sum_{f=0}^{F-1} (10 \log_{10} |D(f)| - \bar{D})^2}, \quad (15)$$

with

$$\bar{D} = \frac{1}{F} \sum_{f=0}^{F-1} 10 \log_{10} |D(f)|.$$

Linear phase deviation is defined as the deviation of the unwrapped phase from a linear fit to its values

$$\Delta = \sqrt{\frac{1}{F} \sum_{f=0}^{F-1} (\theta(f) - \bar{\theta}(f))^2}, \quad (16)$$

where $\bar{\theta}(f)$ is the least squares linear approximation to the phase at frequency bin f .

The results shown here are averaged over 30 different estimation error realizations. The robustness of regularized subband inversion to system estimation errors is now studied as a function of the regularization parameter magnitude. Figure 2 shows the results for $E_m = -10, -20, -30$, and -40 dB.

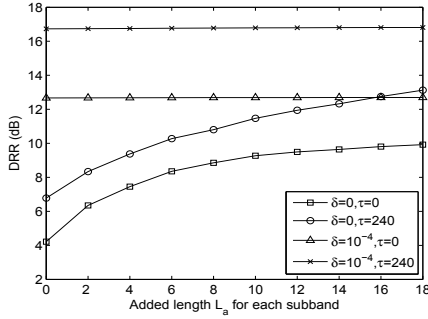


Fig. 5. Performance as a function of added filter length to the reference length for $E_m = -20$ dB. Modeling delay is set at $\tau = 0, 240$, regularization parameter at $\delta = 0, 10^{-4}$.

A regularization parameter, for example, of $\delta = 10^{-4}$ gives the best DRR result for $E_m = -20$ dB, and hereafter, we refer to the value of regularization parameter which gives the best DRR result as the best value with respect to the corresponding system mismatch level.

Figures 3 and 4 shows comparative results between subband and fullband equalization. The dashed line in Fig. 3 represents the averaged DRR of the unprocessed RIRs of the two channels. It can be seen that regularization with the best values improves the performance by about 10 dB in DRR for all system mismatch levels in both the cases of fullband and subband equalization. For system mismatch smaller than -40 dB, the performance of the subband algorithm is constrained by the non-ideal nature of the filter bank design but for system mismatch E_m greater than -40 dB, subband equalization performs at least as well as the fullband. With regularization, the subband algorithm slightly outperforms the fullband one in terms of magnitude and linear phase deviation. Since current blind system estimation algorithms obtain about 6-20 dB normalized projection misalignment (NPM) [17], it can be always expected in practice that $E_m > -40$ dB.

The effects of filter length and modeling delay are shown in Fig. 5. It is seen that, with regularization, increasing the length brings little improvement, but introducing modeling delay significantly improves the performance. Here we only plot the subband case and take $E_m = -20$ dB as an example. Similar results can be seen for the fullband case and other mismatch levels.

According to (12), the subband approach reduces the required computational complexity by a factor of about 125 for our implementation with $M = 2$, $K = 32$ subbands decimated by $N = 24$, and $L = 1600$.

5. CONCLUSION

In this paper, we introduced regularization into the subband multichannel LS equalization algorithm to achieve robust and fast equalization of RIRs when estimation errors are present.

Compared with the fullband algorithm, the computational complexity has been significantly reduced. Simulations have been used to demonstrate the robustness of the algorithm and the results show that it is comparable with the fullband counterpart.

6. REFERENCES

- [1] P. A. Naylor and N. D. Gaubitch, "Speech dereverberation," in *Int. Workshop Acoust. Echo Noise Control*, Eindhoven, Sept 2005.
- [2] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, Issue 1, pp. 165–169, July 1979.
- [3] M. Miyoshi and K. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 145–152, 1988.
- [4] B. D. Radlović, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 8, no. 3, pp. 311–319, May 2000.
- [5] H. Kuttruff, *Room Acoustics*, 4th edition, Taylor & Francis, 2000.
- [6] N. D. Gaubitch, M. R. P. Thomas, and P. A. Naylor, "Subband method for multichannel least squares equalization of room transfer functions," in *WASPAA*, New Paltz, NY, Oct 2007.
- [7] J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *ICASSP*, May 1982, vol. 7, pp. 1858–1861.
- [8] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 5, pp. 882–895, Sept. 2005.
- [9] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, February 2007.
- [10] A. Kataoka K. Furuya, "Robust speech dereverberation using multi-channel blind deconvolution with spectral subtraction," *IEEE Trans. Speech Audio Processing*, vol. 15, no. 5, pp. 1579–1591, July 2007.
- [11] S. Weiss and R. W. Stewart, *On adaptive filtering in oversampled subbands*, Shaker Verlag, 1998.
- [12] J. P. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, "The complex subband decomposition and its application to the decimation of large adaptive filtering problems," *IEEE Trans. Signal Processing*, vol. 50, no. 11, pp. 2730–2743, Nov 2002.
- [13] G. H. Golub and C. F. van Loan, *Matrix computations*, 3rd edition, London: John Hopkins University Press, 1996.
- [14] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise," in *Int. Workshop Acoust. Echo Noise Control*, September 12 - 14, 2006.
- [15] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
- [16] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. Speech Audio Processing*, vol. 7, no. 7, pp. 718–724, Nov. 1999.
- [17] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 174–176, Jul 1998.