APPLICATION OF CHANNEL SHORTENING TO ACOUSTIC CHANNEL EQUALIZATION IN THE PRESENCE OF NOISE AND ESTIMATION ERROR

Mark R. P. Thomas, Nikolay D. Gaubitch and Patrick A. Naylor

Imperial College London
Exhibition Road, London SW7 2AZ, UK
E-mail: {mrt102, ndg, p.naylor}@imperial.ac.uk

ABSTRACT

The inverse-filtering of acoustic impulse responses (AIRs) can be achieved with existing methods provided a good estimate of the channel is available and the observed signals contain little or no noise. Such assumptions are not generally valid in practical scenarios, leading to much interest in the issue of robustness. In particular, channel shortening (CS) techniques have been shown to be more robust to channel estimation error than existing approaches. In this paper we investigate CS using the relaxed multichannel least-squares (RMCLS) algorithm in the presence of both channel error and additive noise. It is shown quantitatively that shortening the acoustic channel to a few ms duration is more robust than attempting to equalize the channel fully, giving better resultant sound quality for dereverberation. A key point of this paper is to provide an explanation for this added robustness in terms of the equalization filter gain. We provide simulation results and results for practical settings using speech recordings and room impulse response measurements from a real acoustic environment.

1. INTRODUCTION

Inverse-filtering of room acoustics is important for applications such as speech dereverberation and listening room compensation. Acoustic impulse responses (AIRs) generally have a nonminimum phase characteristic such that a stable single-channel inverse does not necessarily exist. Single-channel approximations to the inverse problem can be obtained by the method of least squares (LS) but they are of limited use in acoustic channel equalization [1]. When multichannel observations are available, the multiple-input/output inverse theorem (MINT) can provide an exact inverse provided the AIRs are known exactly and do not share any common zeros [2]. In the presence of common zeros, MINT and its multichannel least squares (MCLS) formulation [3] can be shown to completely invert those factors that are not common and perform a LS inversion of those parts with common zeros [4]. Subband and iterative algorithms have also been proposed to reduce computational complexity [5, 6], although the convergence rate of iterative algorithms is often limited due to ill-conditioning of the problem.

Channel shortening (CS) techniques have been extensively developed in the context of digital communications to mitigate inter-symbol and inter-carrier interference, whereby low-order taps are unconstrained in a so-called *relaxation window*. A common framework for CS can be found in [7]. In the equalization of acoustic systems, CS is used to remove audible echoes such that inaudible ones remain, thereby relaxing the design constraints on an equalization filter [8, 9]. A generalization of CS as the shortening/reshaping of an AIR with p-norm optimization aims to exploit additional psychacoustic effects in the design of the equalization filter [10].

Existing inverse-filtering techniques perform well providing the AIR is time-invariant and estimated exactly. However, in many practical scenarios, the AIRs can be disturbed by factors such as source position and temperature that lead to the inversion of a poor estimate of the system [11]. The inverse filter energy is often large, where filter energy is defined as the ℓ_2 norm of the filter coefficients. This amplifies small fluctuations in the AIR, leading to increased reverberation rather than a reduction. Efforts to reduce sensitivity to error have involved constraining the filter energy at the expense of reduced dereverberation [3, 1]. Filter gain can also be reduced by introducing a modelling delay into the equalization filter so as to relax causality constraints. Previous studies have also considered the effect of noise from the recording apparatus and acoustic sources upon inverse-filtering. It has been shown that constraints placed upon the gain of the equalization filter can improve noise robustness in addition to channel estimation error [3, 1]. Channel shortening algorithms have been shown to possess an additional desirable property in that they tend to be more robust to channel estimation errors than existing least-squares techniques [12].

In this paper we consider channel equalization using CS in the presence of both channel estimation error and additive noise. We investigate the robustness of the relaxed multichannel least-squares (RMCLS) algorithm [12] as a function of the relaxation window length and modelling delay. This investigation also provides an insight into an empirical optimum relaxation window length, and shows that RMCLS can be used for practical channel equalization of real-world recordings. The filter gain is also investigated as a function of relaxation window length so as to provide a link between robustness enhancement with CS and explicit gain constraints.

The remainder of this paper is organized as follows. A problem formulation is given in Sec. 2, the least-squares equalization algorithms are reviewed in Sec. 3, performance evaluation is presented in Sec. 4 and conclusions are given in Sec. 5.

2. PROBLEM FORMULATION

Consider a speech signal recorded in a noisy, reverberant environment with an array of microphones. The observed signals at microphone $m \in \{1, 2, \ldots, M\}$ are given by

$$\mathbf{x}_m(n) = \mathbf{H}_m \mathbf{s}(n) + \boldsymbol{\nu}_m(n), \qquad (1)$$

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.

978-1-4577-0693-6/11/$26.00 ©2011 IEEE

where $\mathbf{s}(n) = [s(n)\; s(n-1)\; \dots\; s(n-2L+1)]^T$, $\mathbf{x}_m(n) = [x_m(n)\, x_m(n-1)\; \dots\; x_m(n-L+1)]^T$, $\boldsymbol{\nu}_m(n) = [\nu_m(n)\, \nu_m(n-1)\; \dots\; \nu_m(n-L+1)]^T$ are segments of the speech signal, noisy observation and noise starting at sample n respectively and $\mathbf{H}_m$ denotes the $L \times (2L-1)$ filtering matrix derived from the AIR. Segments are L samples in length. The filtering matrix

$$\mathbf{H}_m = \begin{bmatrix} h_{m,0} & \dots & h_{m,L-1} & \dots & \dots & 0 \\ 0 & h_{m,0} & \dots & h_{m,L-1} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & h_{m,0} & \dots & h_{m,L-1} \end{bmatrix} \tag{2}$$

is derived from the filtering vector $\mathbf{h}_m = [h_{m,0}\; \dots\; h_{m,L-1}]^T$. Impulse responses are assumed to be slowly time-varying such that $\mathbf{h}_m$ is independent of n. By concatenating all M outputs of (1), a system of equations

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) + \boldsymbol{\nu}(n) \tag{3}$$

can be obtained using the following quantities

$$\mathbf{x}(n) = [\mathbf{x}_1^T(n)\, \mathbf{x}_2^T(n)\; \dots\; \mathbf{x}_M^T(n)]^T, \tag{4}$$

$$\mathbf{H} = [\mathbf{H}_1^T\; \mathbf{H}_2^T\; \dots\; \mathbf{H}_M^T]^T, \tag{5}$$

$$\boldsymbol{\nu}(n) = [\boldsymbol{\nu}_1^T(n)\, \boldsymbol{\nu}_2^T(n)\; \dots\; \boldsymbol{\nu}_M^T(n)]^T. \tag{6}$$

The noise signals are assumed to be mutually uncorrelated with the source signal. Equalization filters can be calculated by solving the following system of equations

$$\sum_{m=1}^{M} h_m(j) * g_m(j) = d(j) \quad \text{for } j = 0, \dots, L + L_i - 2, \tag{7}$$

where L_i is the length of g_m and

$$d(j) = \begin{cases} 0 & \text{if } 0 \le j < \tau; \\ 1 & \text{if } j = \tau; \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

represents the target response with delay τ. In matrix form, (7) can be written as

$$\tilde{\mathbf{H}}\mathbf{g} = \mathbf{d}, \tag{9}$$

where $\mathbf{d} = [d(0) \cdots d(L + L_i - 2)]^T$ represents the target response vector and

$$\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1\; \cdots\; \tilde{\mathbf{H}}_M], \tag{10}$$

with $\tilde{\mathbf{H}}_m$ an $(L + L_i - 1) \times L_i$ filtering matrix of $\mathbf{h}_m$,

$$\tilde{\mathbf{H}}_m = \begin{bmatrix} h_{m,0} & 0 & \cdots & 0 \\ h_{m,1} & h_{m,0} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_{m,L-1} & \cdots & \vdots & \vdots \\ 0 & h_{m,L-1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{m,L-1} \end{bmatrix}. \tag{11}$$

Let $\hat{\mathbf{h}}$ be an estimate of $\mathbf{h}$ that contains misalignment error, with corresponding inverse filter $\hat{\mathbf{g}}$ and filtering matrix $\hat{\mathbf{G}}$ defined in a similar fashion to (5). The aim is to estimate the speech signal $\mathbf{s}(n)$ from the noisy, reverberant observations using

$$\hat{\mathbf{s}}(n) = \hat{\mathbf{G}}\mathbf{x}(n). \tag{12}$$

3. LEAST-SQUARES EQUALIZATION

3.1. Least Squares (LS) and MINT

For single channel case, where $M = 1$, (9) is always an over-determined system of equations. The LS solution that minimizes

$$J = \|\tilde{\mathbf{H}}\mathbf{g} - \mathbf{d}\|_2^2, \tag{13}$$

is given by

$$\mathbf{g} = \tilde{\mathbf{H}}^{\dagger}\mathbf{d}, \tag{14}$$

where $\{\cdot\}^{\dagger}$ denotes Moore-Penrose pseudo-inverse [13]. The multichannel least-squares (MCLS) formulation of the MINT algorithm minimizes the cost function in (13) for $M \ge 2$. Exact solution(s) that satisfy (9) exist when the following two conditions are both satisfied [2]:

(C1) $H_m(z^{-1})$, the z-transforms of the multichannel AIRs $\mathbf{h}_m$ do not have any common zeros.

(C2) $L_i \ge L_c$, with $L_c = \lceil \frac{L-1}{M-1} \rceil$ and $\lceil \kappa \rceil$ denotes the smallest integer larger than or equal to κ [14].

3.2. RMCLS

The RMCLS algorithm calculates a CS equalizer for AIRs by relaxation of the target function, which in the case of the MCLS formulation of the MINT algorithm is a unit impulse with $\tau = 0$. We refer to MINT as full equalization. It has been shown that early reflections in the AIR are perceived to reinforce the direct sound and are considered to be beneficial to speech intelligibility. The cost function in (13) is therefore modified with a weighting function,

$$\mathbf{w} = [\underbrace{1\, \dots\, 1}_{\tau}\, \underbrace{1\, 0\, \dots\, 0}_{L_r}\, 1\, \dots\, 1]^T_{(L+L_i-1)\times 1}, \tag{15}$$

where L_r is the length of the *relaxed window*. The modified cost function becomes

$$J = \|\mathbf{W}(\tilde{\mathbf{H}}\mathbf{g} - \mathbf{d})\|_2^2, \tag{16}$$

where $\mathbf{W} = \text{diag}\{\mathbf{w}\}$ is a diagonal weighting matrix with $\mathbf{w} = [w_0\; \dots\; w_{L+L_i-2}]^T$, $w_i \neq 0\; \forall\, i$, that is minimized by

$$\mathbf{g} = (\mathbf{W}\tilde{\mathbf{H}})^{\dagger}\mathbf{W}\mathbf{d}. \tag{17}$$

The weighting function in (15) contains entries exactly equal to 0 within the relaxed window such that in the minimization of (16), the samples in the relaxed window are unconstrained. The first entry in the relaxed window is set to 1 to avoid the trivial solution.

3.3. Robustness Issues

Experimental studies have revealed that existing least-squares equalization algorithms are particularly sensitive to (a) noise and (b) channel estimation error [3]. It has been shown that CS algorithms are more robust to (a) [12] than existing LS algorithms and that an additional constraint upon filter gain significantly improves robustness to both (a) and (b) [3]. These findings lead to some outstanding questions regarding the robustness of CS equalization that are investigated in the following section:

1. The performance of a CS equalizer as a function of both noise and channel estimation error.
2. Whether there exists an empirical optimum L_r for a given condition in 1.
3. The relationship between and L_r and filter gain.
4. The feasibility of CS equalization for real measurements.

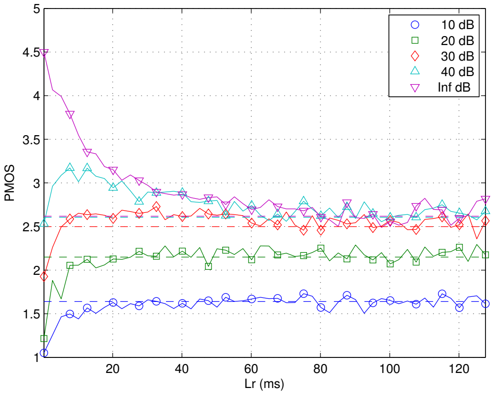

Figure 1: PESQ PMOS results with fixed channel misalignment and variable SNR as a function of L_r. Solid line: processed, dashed line: unprocessed.

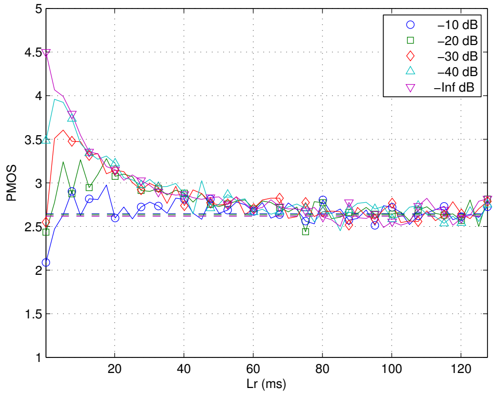

Figure 2: PESQ PMOS results with fixed SNR and variable channel misalignment as a function of L_r. Solid line: processed, dashed line: unprocessed.

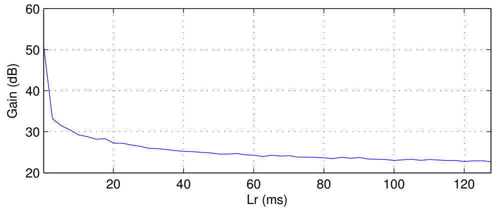

Figure 3: Mean equalization filter gain as a function of L_r.

4. EVALUATION

The performance of RMCLS was evaluated under controlled SNR and channel misalignment as a function of relaxation window length L_r. In the case $L_r = 1$, RMCLS is identical to the MCLS implementation of MINT. In experiment 1, simulated results aim to provide insight into an empirical optimum L_r for the condition under test. In experiment 2, a real-world experiment was conducted to provide audio examples.

4.1. Experimental Setup 1

A room measuring $5 \times 4 \times 4$ m was simulated using the source-image method [15] with sampling frequency $f_s = 8$ kHz, $M = 2$, $L = 1024$. Gaussian distributed channel errors were generated proportional to the filter taps such that the misalignment between the true AIR, $\mathbf{h}$, and corrupted AIR, $\hat{\mathbf{h}}$, was $[-\infty, -40, -30, -20$ and $-10]$ dB. Equalization filters $\hat{\mathbf{g}}$, with $L_i = L_c = 1023$, were calculated from $\hat{\mathbf{h}}$ with delay $\tau = 0$. Noise was added at the receiver to obtain a received SNR in $[10, 20, 30, 40, \infty]$ dB. For each combination of channel error and additive noise level, a source was placed in 20 random locations 2 m from a pair of receiving microphones. A clean speech signal from the SAM database [16] of duration 10 s was produced at the source and reverberant observations calculated with (3). The speech signal was estimated from the noisy observations with (12). The clean and processed speech signals were analysed using the objective speech quality measure ITU-T P.862 (PESQ) [17] to estimate perceptual speech quality as a predicted mean opinion score (PMOS) in the range 1–4.5. The averaged results are summarized in Figs. 1 and 2, where solid and dashed lines are the PESQ score of the processed and unprocessed speech signals respectively. The average filter gain $E\{\|\mathbf{g}(k)\|_2\}$, where k enumerates the solution and $E\{\cdot\}$ denotes mathematical expectation, is plotted in Fig. 3.

4.2. Experimental Setup 2

Two DPA 4060 microphones spaced by 16 cm were placed in a $5.14 \times 3.16 \times 3$ m room with reverberation time (T_{60}) of 170 ms. A Genelec 8030a loudspeaker was placed 1.2 m away and channel estimates, $L = 1366$ at $f_s = 8$ kHz, were found with MLS [18]; higher sampling rates require optimization of the software implementation due to a large memory footprint. A speech signal from the SAM database [16] was recorded using the same configuration. Inverse filters of length $L_i = L_c = 1365$ were calculated using the MLS-derived channel estimates for a range of L_r and applied to the recorded signals in a similar way to Experiment 1. It was expected that source position and temperature variation in the room contributed to the overall channel misalignment [3]. Microphone self-noise and acoustic noise from air ducts etc. contributed to the noise. PESQ results are shown in Table 1. [1]

4.3. Discussion

The results in Figs. 1 and 2 represent the full range of channel shortening values. When $L_r = 1$, RMCLS is equivalent to the MCLS implementation of MINT; when $L_r = L_i$, the equalizer taps take random values as they are entirely unconstrained.

In the case where SNR=∞ dB, misalignment is $-\infty$ dB, and $L_r = 1$, PMOS takes the maximum value of 4.5 as the AIR is perfectly inverted. The introduction of CS in this case is detrimental to the quality of the processed signal. The addition of noise or misalignment leads to an increase in the empirical optimum L_r, where noise has a more significant influence over the speech quality. Providing the SNR is greater than $\sim$ 30 dB, channel equalization can benefit significantly from CS; below 30 dB CS does not give improved results but are generally no lower than the unprocessed values. In all cases, PMOS scores reach a maximum and converge to

[1] Examples of clean, unprocessed and processed speech can be found at http://www.commsp.ee.ic.ac.uk/∼mrt102/projects/samples.html.

Table 1: PESQ PMOS results for a real-world measurement.

Unproc.	L_r=1	L_r=50	L_r=100	L_r=200	L_r=400
	(0 ms)	(6.25 ms)	(12.5 ms)	(25 ms)	(50 ms)
2.6	2.2	3.0	3.1	2.9	2.6

the unprocessed PMOS with increasing L_r. As a rule-of-thumb, a relaxation window in the order of 10 ms provides better results than attempting full equalization.

The inverse-filtering approach in [3] suggests that increased robustness to channel error and noise is achieved by reducing the filter gain. The results in Fig. 3 show the mean filter gain for Experiment 1 as a function of L_r and reveal that it decreases almost monotonically with increased channel length, with the most pronounced effect within the first 10 ms. This is largely consistent with the PMOS scores in this experiment. The advantage of CS over explicit gain constraints is that CS forces errors to lie only within the relaxation window where they are less perceivable. It was also suggested in [3] that the introduction of a modelling delay, such that $\tau > 0$, reduces gain by relaxing causality constraints. The same experiment was conducted for $\tau = L/2$, giving marginally improved PMOS scores but with similar characteristics as those in Figs. 1 and 2.

The audio results of Experiment 2 are consistent with those of Experiment 1. When $L_r = 1$, the audio samples exhibit a much greater noise floor than the unprocessed signals and the reverberation time appears to be increased. This is reflected in the low PMOS score of 2.2. As in Experiment 1, a small amount of channel shortening significantly improves the results. A high value of L_r introduces noticeable spectral distortion caused by the unconstrained filtering within the relaxation window. PMOS peaks at 3.1 when $L_r = 100$. Perceptually the noise drops, dereverberation becomes appreciable and the talker appears closer to the microphone.

5. CONCLUSIONS

An experimental study into the robustness of the RMCLS channel shortening algorithm for acoustic channel equalization has been conducted. Building upon the known result that CS improves robustness to channel misalignment, simulated and real-world experiments considered the additional problem of measurement noise. Objective speech quality measurements with PESQ reveal that channel inversion is more sensitive to noise than channel misalignment, and that an empirical optimum exists that is dependent upon the level of distortion. It was also shown that channel shortening reduces the equalizer gain, which is already known to increase the robustness of a channel equalizer. These results motivate an analytical study into the performance of channel shortening algorithms with noise and channel misalignment.

6. REFERENCES

[1] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer, 2010.

[2] M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[3] T. Hikichi, M. Delcroix, and M. Miyoshi, “Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations,” *EURASIP Journal on Applied Signal Processing*, vol. 2007, 2007.

[4] W. Zhang, A. W. H. Khong, and P. A. Naylor, “Adaptive inverse filtering of room acoustics,” in *Proc. Asilomar Conf. on Signals, Systems and Computers*, 2008.

[5] N. D. Gaubitch and P. A. Naylor, “Equalization of multichannel acoustic systems in oversampled subbands,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1061–1070, Aug. 2009.

[6] K. Furuya and Y. Kaneda, “Two-channel blind deconvolution for non-minimum phase impulse responses,” in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 1997, pp. 1315–1318.

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

[9] M. Kallinger and A. Mertins, “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- and p-norm optimization,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 249–259, Feb. 2010.

[11] F. Talantzis and D. B. Ward, “Robustness of multichannel equalization in an acoustic reverberant environment,” *J. Acoust. Soc. Am.*, vol. 114, no. 2, pp. 833–841, 2003.

[12] W. Zhang, E. A. P. Habets, and P. A. Naylor, “On the use of channel shortening in multichannel acoustic system equalization,” in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.

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

[14] 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.

[15] P. M. Peterson, “Simulating the response of multiple microphones to a single acoustic source in a reverberant room,” *J. Acoust. Soc. Am.*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.

[16] D. Chan, A. Fourcin, D. Gibbon, B. Granstrom, M. Huckvale, G. Kokkinakis, K. Kvale, L. Lamel, B. Lindberg, A. Moreno, J. Mouropoulos, F. Senia, I. Trancoso, C. Veld, and J. Zeiliger, “EUROM - a spoken language resource for the EU,” in *Proc. European Conf. on Speech Communication and Technology*, Sept. 1995, pp. 867–870.

[17] ITU-T, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*, International Telecommunications Union (ITU-T) Recommendation P.862, Feb. 2001.

[18] J. Vanderkooy, “Aspects of MLS measuring systems,” *Journal Audio Eng. Soc.*, vol. 42, pp. 219–231, 1994.