AN EFFICIENT MULTICHANNEL EQUALIZATION ALGORITHM FOR AUDIO APPLICATIONS

Jibran Yousafzai and Zoran Cvetković

King's College London
Strand, London, WC2R 2LS, UK
$[jibran.yousafzai, zoran.cvetkovic]@kcl.ac.uk$

ABSTRACT

The challenge of multichannel equalization for audio applications lies in the physical properties of the underlying multi-input/multi-output (MIMO) linear time-invariant systems which are generally non-minimum phase and exhibit extremely long impulse responses, thereby imposing a considerable computational burden on the equalization task particularly when iterative solutions are sought. In this paper we propose a computationally efficient non-iterative multichannel equalization algorithm. The proposed algorithm is based on the Fast Fourier Transform (FFT) and allows for faster and considerably more accurate inversion of MIMO systems compared to traditional deconvolution algorithms and adaptive solutions. We address the accuracy and limitations of the proposed algorithm and present simulation results illustrating its performance.

Index Terms— MIMO systems, multichannel equalization, dereverberation, audio systems.

1. INTRODUCTION

Multichannel equalization plays an important role in sound reproduction systems. It is the central problem in audio applications which require dereverberation of acoustic sources and cross-talk cancellation for soundfield rendering.

The goal of dereverberation is to extract dry audio signals by post-processing the convolutive mixture of microphone signals in order to remove the effect of room acoustics. On the other hand, cross-talk cancellation attempts to pre-process the reverberant microphone signals so that on playback, the listener perceives an unmodified stereophonic signal. These techniques aim to compensate the deficiencies of the transduction chain for the reproduction of a spatially coherent soundfield [1]. One particular application which was the main motivation for the work presented in this paper is the separation of direct and diffuse components of the individual audio signals for the reconstruction of perceptual soundfield [3, 4]. In either case, the underlying mathematical problem to be solved involves the inversion of an LTI MIMO system. The particular challenge of this problem for the aforementioned applications is that the corresponding MIMO systems have non-minimum phase characteristics and feature very long impulse responses.

The non-blind and semi-blind iterative source separation schemes studied previously [1, 2] require very long training sequences for convergence of the inverse filters and therefore place a considerable computational burden on the system. A non-iterative dereverberation algorithm was presented in [5]. This algorithm was however restricted to one audio source. In this paper, we propose a non-iterative FFT-based solution to multichannel equalization which is able to invert MIMO systems corresponding to multiple sound sources. The inversion of acoustics is performed in the frequency domain under the assumption that the impulse responses of the performance venue between the sound sources and the microphones are known. This is a minor restriction since the impulse responses can be measured [6, 7, 8]. The proposed method is several orders of magnitude faster and more precise than the conventional iterative solutions. For simplicity, we restrict our discussion to a post-equalization structure for dereverberation of audio signals as shown in Figure 1.

In Section 2, we set a mathematical formulation of the problem and present the proposed solution. The simulations and results of error performance of the equalizer are reported in Section 3. Finally, Section 4 draws some conclusions.

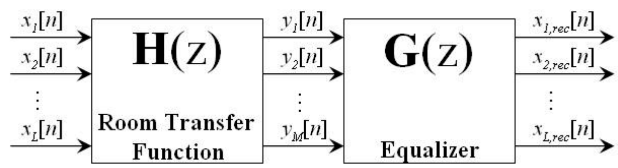


Fig. 1. *Inversion of a MIMO system $\mathbf{H}(z)$ by an equalizer $\mathbf{G}(z)$.*

2. MULTICHANNEL EQUALIZATION USING FFT

Consider L instruments playing in an acoustic space and M microphones recording the soundfield . The signal captured by m^{th} microphone is given by

$$Y_m(z) = \sum_{l=1}^{L} H_{lm}(z) X_l(z) \tag{1}$$

where $X_l(z)$ is the signal of the l^{th} instrument and $H_{lm}(z)$ is the transfer function of the space between l^{th} instrument and m^{th} microphone. The problem addressed in this paper is to reconstruct (dereverberate) signals $X_1(z), \ldots, X_L(z)$ from their convolutive mixtures $Y_1(z), \ldots, Y_M(z)$. In matrix notation, the microphone signals are given by

$$\mathbf{Y}(z) = \mathbf{H}(z)\mathbf{X}(z)$$

where

$$\mathbf{Y}(z) = [Y_1(z), \ldots, Y_M(z)]^T \ , \ \mathbf{X}(z) = [X_1(z), \ldots, X_L(z)]^T \ ,$$

and

$$\mathbf{H}(z) = \begin{bmatrix} H_{11}(z) & \dots & H_{L1}(z) \\ \vdots & \ddots & \vdots \\ H_{1M}(z) & \dots & H_{LM}(z) \end{bmatrix} . \tag{2}$$

The dereverberation requires finding a matrix of equalization filters,

$$\mathbf{G}(z) = \begin{bmatrix} G_{11}(z) & \dots & G_{1M}(z) \\ \vdots & \ddots & \vdots \\ G_{L1}(z) & \dots & G_{LM}(z) \end{bmatrix} ,$$

such that $\mathbf{M}(z) = \mathbf{G}(z)\mathbf{H}(z)$, the transfer function of the cascade of the acoustic space and the equalizer $\mathbf{G}(z)$, is a pure delay,

$$\mathbf{M}(z) = \mathbf{G}(z)\mathbf{H}(z) \equiv z^{-\Delta}\mathbf{I}_{LxL}(z) . \tag{3}$$

A necessary and sufficient condition for the existence of such a matrix of stable filters is that $\mathbf{H}(z)$ is of full-rank everywhere on the unit circle. The minimum norm solution for $\mathbf{G}(z)$ is then provided by the left pseudo-inverse of $\mathbf{H}(z)$,

$$\mathbf{G}(z) = \left(\mathbf{H}^T(z^{-1})\mathbf{H}(z)\right)^{-1}\mathbf{H}^T(z^{-1}) \tag{4}$$

Exact computation of the pseudoinverse of $\mathbf{H}(z)$ is numerically prohibitive, since its entries are polynomials of very high orders, *e.g.* around $44,000$ for 1s reverberation time at 44.1kHz sampling. Furthermore, $\mathbf{G}(z)$ will be non-causal and will result in IIR filters if $\left|\mathbf{H}^T(z^{-1})\mathbf{H}(z)\right|$ is not a pure delay. Below, we propose a numerically efficient algorithm to find an FIR approximation of the left pseudoinverse of $\mathbf{H}(z)$.

Let

$$\mathbf{B}(z) = \begin{bmatrix} B_{11}(z) & \dots & B_{1L}(z) \\ \vdots & \ddots & \vdots \\ B_{L1}(z) & \dots & B_{LL}(z) \end{bmatrix} = \mathbf{H}^T(z^{-1})\mathbf{H}(z) . \tag{5}$$

Then

$$\mathbf{G}(z) = \mathbf{B}^{-1}(z)\mathbf{H}^T(z^{-1}) \tag{6}$$

and

$$\mathbf{B}^{-1}(z) = \frac{\begin{bmatrix} CofB_{11}(z) & \dots & CofB_{1L}(z) \\ \vdots & \ddots & \vdots \\ CofB_{L1}(z) & \dots & CofB_{LL}(z) \end{bmatrix}^T}{D(z)} \tag{7}$$

where

$$D(z) = \left|\mathbf{B}(z)\right| = Determinant \quad of \quad \mathbf{B}(z)$$

and

$$CofB_{ij}(z) = (-1)^{i+j}\left|B_{kn}(z)\right|, \quad k \neq i, n \neq j$$

Since $CofB_{ij}(z)$ and $D(z)$ are polynomials in z, it should be noted that if we try to invert the matrix $\mathbf{B}(z)$ directly, the inverse matrix $\mathbf{B}^{-1}(z)$ will result in IIR filters. This, of course, is not an ideal solution. However, we can use this direct matrix inversion approach to approximate the inverse IIR filters with FIR filters. The FIR approximation to $\mathbf{B}^{-1}(z)$ are obtained by dividing the N-point DFT of the corresponding cofactors, $CofB_{ij}(z), i = 1, \dots, L, j = 1, \dots, L$, by the N-point DFT of $D(z)$.

$$\mathbf{B}^{-1}(e^{j\frac{2\pi}{N}k}) = \frac{\begin{bmatrix} CofB_{11}(e^{j\frac{2\pi}{N}k}) & \dots & CofB_{1L}(e^{j\frac{2\pi}{N}k}) \\ \vdots & \ddots & \vdots \\ CofB_{L1}(e^{j\frac{2\pi}{N}k}) & \dots & CofB_{LL}(e^{j\frac{2\pi}{N}k}) \end{bmatrix}^T}{D(e^{j\frac{2\pi}{N}k})} \tag{8}$$

$k = 0, 1, \dots, N-1$. Then, the N-point inverse discrete Fourier transform of (8) results in an FIR approximation of the matrix $\mathbf{B}^{-1}(z)$. Finally, the equalizer $\mathbf{G}(z)$ can be obtained from (6). It should be noted that the size of the FFT (N) must be greater than or equal to the length of $D(z)$. The minimum size of the FFT, therefore, is given by

$$FFTSize_{Min} = L_d = 2L(L_h - 1) + 1 \tag{9}$$

where L_h is the length of room impulse response and L_d is the length of $D(z)$. Accordingly, the minimum length that the inverse filters can have is given by

$$L_{g,Min} = L_d + L_h - 1 = 2L(L_h - 1) + L_h . \tag{10}$$

This algorithm computes the coefficients of IIR filters $G_{lm}(z)$ by finding the inverse Fourier transform using finitely many transform samples. This discretization of the Fourier transform causes time-aliasing of $\mathbf{B}^{-1}(z)$ which is reduced as the size of FFT is increased. The impact of the size of the FFT on the relative error of the equalized signals is reported in the next section.

3. DESIGN EVALUATION AND RESULTS

Having previously described the mathematical design, this section presents the evaluation of the equalization algorithm described in Section 2. For comparison, a semi-blind adaptive multichannel equalization algorithm presented in [2] was also implemented. This method uses a multichannel normalized least mean square (M-NLMS) algorithm for the gradient estimation and the update of the adaptive inverse filters.

A quantitative performance measure used to evaluate these algorithms is the Relative Error given by

$$RelativeError = \frac{MSE}{Energy_{Average}} = \frac{\sum_n \left|x[n] - x_{rec}[n]\right|^2}{\sum_n \left|x[n]\right|^2} . \tag{11}$$

Impulse responses, $H_{km}(z)$, were generated for hypothetical rectangular auditoria using the method of images described in [9]. Since the adaptive equalizer requires very long time for training, we use relatively short impulse responses in the numerical experiments so as to compare both algorithms. However, the algorithm proposed in this paper can effectively equalize longer impulse responses as well.

Here we present results to establish post-equalization of audio signals using both algorithms for the following two cases: $L = 2$, $M = 5$ and $L = 3$, $M = 5$. Dry test signals used were: jazz trumpet and saxophone in the $L = 2$ case, and electric jazz guitar, jazz trumpet, and saxophone in the $L = 3$ case. All test signals were 23s high quality audio files, sampled at 44.1kHz, and recorded with a close microphone technique to minimize early reflections and reverberation. The quantitative results and impulse responses of the equalized system for the two scenarios are presented in Tables $1-4$ and Figure 2, respectively. In both cases the size of the FFT used in the proposed algorithm was set to be twice the minimum size given in (9). In the case of two sources, the adaptive algorithm was trained using a sequence of $400,000$ samples, while in the case of three sources, the training sequence was $600,000$ samples long. We can observe from Tables $1-4$ that the proposed FFT-based algorithms attains a $40-50$dB higher accuracy than the adaptive algorithm in the case of two sound sources, and over 60dB higher accuracy in the case of three sources. This improvement is paid by considerably longer filters of the FFT-based equalizer compared to the adaptive algorithm. The number of coefficients in the filters of the adaptive equalizer was set to be equal to length of the room impulse response, since we found that longer or shorter filters were yielding less accurate results. In terms of numerical complexity, the adaptive algorithm requires long training sequences for the adaptive filters to converge and is, therefore, computationally considerably less efficient than the proposed method.

Table 1. *Quantitative results of multichannel equalization using the adaptive equalizer in the case of $L = 2$ source signals and $M = 5$ microphones. Each column corresponds to an individual source signal. L_g - the length of the equalizer filters is set to be equal to L_h - the length of the room impulse responses.*

$L_h -$	1700	
$Training\ Sequence$	400000 $samples$	
L_g	1700	
$Energy_{Average}$	-33.8dB	-26.8dB
MSE	-47.5dB	-43.9dB
$Relative\ Error$	-13.7dB	-17.1dB

Table 2. *Quantitative results of multichannel equalization using the FFT-based equalizer in the case of $L = 2$ source signals and $M = 5$ microphones. Each column corresponds to an individual source signal. L_g - the length of the equalizer filters. L_h - the length of the room impulse responses.*

L_h	1700	
$FFTSize$	$2 * FFTSize_{Min}$	
L_g	15302	
$Energy_{Average}$	-33.8dB	-26.8dB
MSE	-98.2dB	-82.1dB
$Relative\ Error$	-64.4dB	-55.3dB

Table 3. *Quantitative results of multichannel equalization using the adaptive equalizer in the case of $L = 3$ source signals and $M = 5$ microphones. Each column corresponds to an individual source signal. L_g - the length of the equalizer filters is set to be equal to L_h - the length of the room impulse responses.*

L_h	1700		
$Training\ Sequence$	600000 $samples$		
L_g	1700		
$Energy_{Average}$	-28.2dB	-33.8dB	-26.8dB
MSE	-41.2dB	-45.7dB	-36.9dB
$Relative\ Error$	-13dB	-11.9dB	-10.1dB

Table 4. *Quantitative results of multichannel equalization using the FFT-based equalizer in the case of $L = 3$ source signals and $M = 5$ microphones. Each column corresponds to an individual source signal. L_g - the length of the equalizer filters, L_h - the length of the room impulse responses.*

L_h	1700		
$FFTSize$	$2 * FFTSize_{Min}$		
L_g	22102		
$Energy_{Average}$	-28.2dB	-33.8dB	-26.8dB
MSE	-111.9dB	-109.1dB	-98.8dB
$Relative\ Error$	-83.7dB	-75.3dB	-72dB

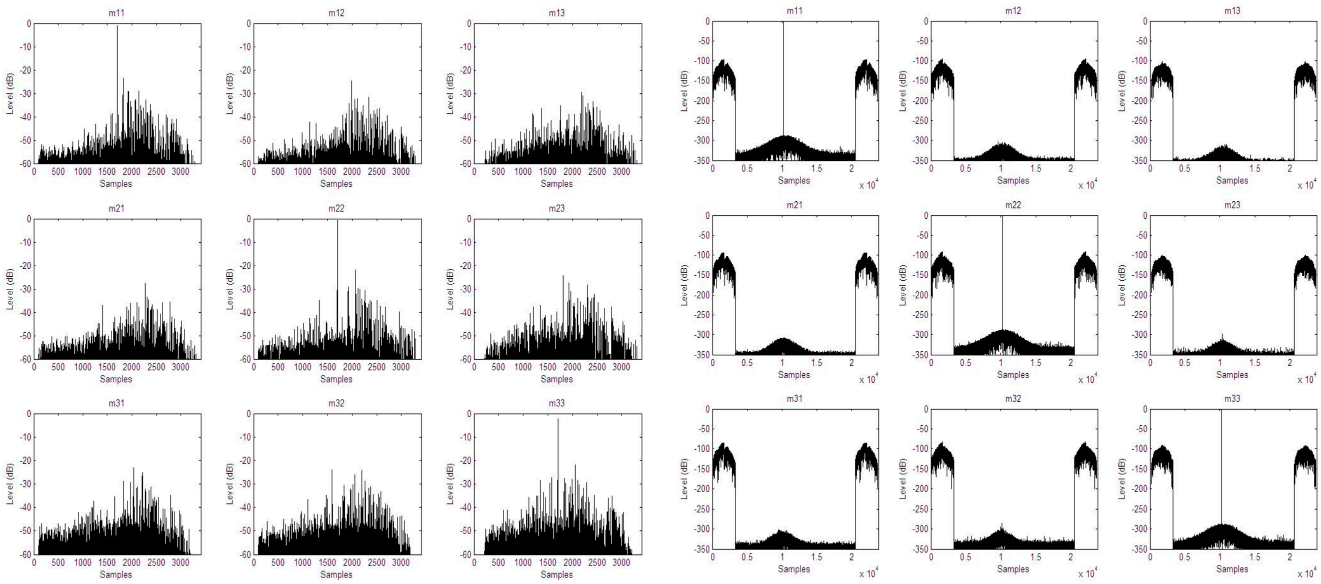
Table 5. *$FFTSize$ vs. $RelativeError$ for the case of $L = 2$ instruments and $M = 5$ microphones. Each column corresponds to an individual instrument.*

$FFTSize$	$RelativeError$	
$2 * FFTSize_{Min}$	-64.4dB	-55.3dB
$4 * FFTSize_{Min}$	-96.9dB	-89.1dB
$8 * FFTSize_{Min}$	-130.2dB	-123.1dB
$10 * FFTSize_{Min}$	-176.5dB	-170.1dB

Impulse responses of the equalized system for the case of $L = 3$ instruments and equalization using the adaptive algorithm and the algorithm proposed here are shown in Figure 2. The maximal error of the FFT-based algorithm is over 60dB below the maximal error of the adaptive algorithm. In Figure 2(b), one can observe error "bumps" at the tails of impulse responses of the equalized system in the case of the FFT-based algorithm. These are a result of the time aliasing due to the discretization of the Fourier transform, as discussed in the previous section. Nevertheless, even at these bumps the error is still around -100dB.

Finally we investigated the impact of the size of the FFT on the equalization accuracy. Tables $5-6$ illustrate the effect of the FFT size on the relative error of dereverberation for the same mixtures of $L = 2$ and $L = 3$ signals, respectively, which were used for experiments shown in Tables 2 and Table 4. An increase in the size of the FFT reduces the time aliasing of the inverse filters, hence decreasing the relative error accordingly. Results shown in Tables $5-6$ suggest that in this way the error could be made arbitrarily small. But increasing the size of the FFT in turn increases the length of the inverse filters. Therefore, the size of the FFT should be kept moderate enough such that the inverse filters are not very long and the relative error is small enough so that the difference between the original dry source signals and the reconstructed signals is below the level of human hearing.

The FFT-based equalizer has its limitations. If the condition $L < M$ is not satisfied, $D(z)$ is very close to zero because the matrix $\mathbf{H}(z)$ is not well-conditioned at all frequencies. Hence, accurate inversion of the system could not be achieved regardless of the FFT size. Therefore, a restriction of this algorithm is that the number of sound sources is less than the number of microphones capturing the auditory scene.



(a) (b)

Fig. 2. *Impulse response of the equalized system in the case of $L = 3$ sound sources and $M = 5$ microphones. (a) Adaptive equalizer (b) FFT-based equalizer.*

Table 6. *$FFTSize$ vs. $RelativeError$ for the case of $L = 3$ instruments and $M = 5$ microphones. Each column corresponds to an individual instrument.*

$FFTSize$	$RelativeError$		
$2 * FFTSize_{Min}$	-83.7dB	-75.3dB	-72dB
$4 * FFTSize_{Min}$	-122.1dB	-121.3dB	-119.9dB
$8 * FFTSize_{Min}$	-180.9dB	-179.8dB	-170.8dB
$10 * FFTSize_{Min}$	-227.5dB	-220.9dB	-212.3dB

4. CONCLUSION

A fast FFT-based multichannel equalization algorithm for audio applications has been proposed which enables the inversion of long multiple acoustic impulse responses. The performance of the algorithm has been evaluated in numerical experiments. It is important to emphasize that this algorithm is effective for any number of sound sources provided the number of microphones is greater than the number of sound sources in the performance venue. A drawback of this approach is that the inverse filters are quite long. On the other hand, this scheme provides significantly higher accuracy and is computationally much more efficient than the iterative equalization algorithms [1, 2].

5. REFERENCES

[1] M. Guillaume, Y. Grenier and G. Richard, "Iterative Algorithms for Multichannel Equalization in Sound Reproduction Systems," *ICASSP'05*, Philadelphia, USA, March 2005.

[2] S. Weiss, G.W. Rice and R.W. Stewart, "Multichannel Equalization in Subbands," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 203-206, New Paltz, New York, October 1999.

[3] J.D. Johnston and Y.H. Lam, "Perceptual soundfield reconstruction," *109th AES Convention*, paper No. 5202, September 2000.

[4] G.L. Rosen and J.D. Johnston, "Automatic Speaker Directivity Control for Soundfield Reconstruction," *19th International AES Conference*, Schloss Elmau, Germany, June 2001.

[5] B.S. Olswang and Z. Cvetković, "Separation of Audio Signals into Direct and Diffuse Soundfields for Surround Sound," *ICASSP'06*, Toulouse, France, May 2006.

[6] M.R. Schroeder, "Integrated-impulse method measuring sound decay without using impulses," *Journal of the Acoustic Society of America*, Vol. 66, No. 2, pp. 497-500, 1979.

[7] D.D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *Journal of the Acoustic Society of America*, Vol. 37, No. 6, pp. 419–444, 1989.

[8] G.B. Stan, J.J. Embrechts and D. Archambeau, "Comparison of different impulse response measurement techniques,"*Journal of the Acoustic Society of America*, Vol. 50, No. 4, pp. 249–262, 2002.

[9] J.B. Allen and D.A. Berkeley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustic Society of America*, Vol. 65, No. 4, pp. 943–950, April 1979.