Equalization of Multichannel Acoustic Systems in Oversampled Subbands

Nikolay D. Gaubitch, *Member, IEEE,* Patrick A. Naylor*, *Senior Member, IEEE,*

***Abstract*— Equalization of room transfer functions (RTFs) is an important topic with several applications in acoustic signal processing. RTFs are often modelled as finite impulse response filters, characterized by orders of thousands of taps and non-minimum phase. In practice, only approximate estimates of the actual RTFs are available due to measurement noise, limited estimation accuracy and temporal variation of source-receiver position. These issues make equalization a difficult problem. In this paper, we discuss multichannel equalization with focus on inexact RTF estimates. We present a multichannel method for the equalization filter design utilizing decimated and oversampled subbands, where the full-band acoustic impulse response is decomposed into equivalent subband filters prior to equalization. This technique is not only more computationally efficient but also more robust to impulse response inaccuracies compared with the full-band counterpart.**

***Index Terms*—**

EDICS Category: a) AUD-SSEN, b) AUD-LMAP

I. Introduction

EQUALIZATION of room transfer functions (RTFs) is an important research topic with several applications in acoustic signal processing, including speech dereverberation [1] and sound reproduction [2]. Although, in theory, exact equalization is possible when multiple observations are available [3], there are many obstacles for practical application of RTF equalization algorithms.

Consider the L-tap room impulse response of the acoustic path between a source and the mth microphone in an M-element microphone array, $\mathbf{h}_m = [h_{m,0}\ h_{m,1}\ \dots\ h_{m,L-1}]$, with a z-transform $H_m(z)$ constituting the RTF. The objective of equalization is to apply an inverse system with transfer function $G_m(z)$ such that

$$H_m(z)G_m(z) = \kappa z^{-\tau}, \quad m = 1, 2, \dots, M \tag{1}$$

where τ and κ are arbitrary delay and scale factors respectively. Equivalently, considering the L_i tap impulse response of $G_m(z)$, $\mathbf{g}_m = [g_{m,0}\ g_{m,1}\ \dots\ g_{m,L_\mathrm{i}-1}]^T$, (1) can be written in the time domain as

$$\mathbf{H}_m\mathbf{g}_m = \mathbf{d}, \tag{2}$$

Manuscript received May 4, 2007. This work was supported by the Engineering and Physical Sciences Research Council, UK. Grant number GR/S66954.

Nikolay D. Gaubitch and Patrick A. Naylor are with Imperial College London.

where

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

is a $(L + L_\mathrm{i} - 1) \times L_\mathrm{i}$ convolution matrix, and

$$\mathbf{d} = [\underbrace{0\ \dots\ 0}_{\tau}\ \kappa\ 0\ \dots\ 0]^T$$

is the $(L + L_\mathrm{i} - 1) \times 1$ vector with the impulse response of the equalized RTF.

The problem of equalization is to find $G_m(z)$. When $H_m(z)$ is a minimum phase system, a stable inverse filter can be found by replacing the zeros of $H_m(z)$ with poles [4]

$$G_m(z) = \frac{1}{H_m(z)}. \tag{3}$$

However, RTF equalization is not that straightforward in practice because:

(i) RTFs are non-minimum phase in general [5] and hence (3) does not give a stable causal solution for $G_m(z)$;
(ii) the average difference between maxima and minima in RTFs are in excess of 10 dB [6]–[8] and therefore RTFs typically contain spectral nulls that, after equalization, give strong peaks in the spectrum causing narrow band noise amplification;
(iii) equalization filters designed from inaccurate estimates of $H_m(z)$ will cause distortion in the equalized signal [8];
(iv) the length L of $\mathbf{h}_m$ at a sampling frequency f_s is related to the reverberation time, T_{60}, in a room by $L \propto f_s T_{60}$ and can be several thousand taps in length [6].

Several alternative approaches, both for single and for multiple microphones, have been proposed to address these issues. There are two common methods for single channel equalization: single channel least squares (SCLS) and homomorphic equalization [9]. SCLS equalization filters are designed by minimizing an error formed from (2) as [9], [10]

$$\hat{\mathbf{g}}_m = \arg\min_{\mathbf{g}_m} \|\mathbf{H}_m\mathbf{g}_m - \mathbf{d}\|^2, \tag{4}$$

where $\|\cdot\|$ denotes Euclidean distance. In homomorphic inverse filtering [9], [11]–[13], the RTF is decomposed into

minimum phase and all-pass components. An exact inverse can be found for the minimum phase component with (3), while the all-pass component can be equalized, for example, using a matched filter [12]. Equalizing only the magnitude was considered in [5], [12], but was found to result in audible residual echoes. In a comparative study between these two techniques, Mourjopoulos [9] concluded that SCLS, although sometimes less accurate than homomorphic inversion, is more efficient in practice.

Single channel methods typically result in large processing delay, which is problematic for many communications applications, extremely long and non-causal inverse filters, and provide only approximate equalization [3]. Due to the approximate nature of these inverse filters, they are less sensitive to noise and inexact RTF estimates [1]. Inherently, SCLS inverse filters only partially equalize deep spectral nulls, which can be advantageous in avoiding problems due to points (ii) and (iii) above.

In the multichannel case, the non-minimum phase problem is eliminated and exact inversion can be achieved using the Bezout's theorem [3], [14]: given a set of M RTFs, $H_m(z)$, and assuming that these do not have any common zeros, a set of filters, $G_m(z)$, can be found such that [3], [14]

$$\sum_{m=1}^{M} H_m(z)G_m(z) = 1. \tag{5}$$

MINT [3] was the first multichannel equalization method based on (5). Adaptive versions have also been considered [2]. Unlike single channel equalization filters, the length of the multichannel equalization filters is of similar order as the length of the room impulse responses and there is no processing delay [3], [14]. However, it has been observed that exact equalization is of limited value in practice, when the RTF estimates contains even moderate errors [1], [8].

Various alternatives have been proposed for improving robustness to RTF inaccuracies. Bharitkar et al [15] use spatially averaged RTFs for the design of the equalization filter. In [16], the authors modify the desired signal in the multichannel inverse filter design, such that the late reverberation is equalized while the early reflections are preserved. Haneda et al [17], [18] form an IIR filter by decomposing the RTFs into common acoustical poles and non-common zeros. Mourjopoulos [10] uses an AR model of the RTFs rather than the all-zero model in order to reduce the filter order. The AR model of RTFs is also exploited by Hopgood and Rayner in a single channel subband equalization approach [19]. Hikichi et al [20], [21] introduce regularized multichannel equalization which adds robustness to noise and RTF fluctuations.

In this paper, we propose a new method for equalization filter design. Given a set of multichannel RTFs, we decompose the RTFs into their subband equivalent filters. These are then used to design the subband equalization filters and the equalization is performed in each subband before a full-band equalized signal is reconstructed. It is shown that this approach not only reduces the computational load, but also reduces the sensitivity to estimation errors and the effect of measurement noise in the RTFs. An important result is that this method accommodates multichannel equalization of large order systems, taking advantage of the shorter length of multichannel equalization filters and low sensitivity to RTF inaccuracies.

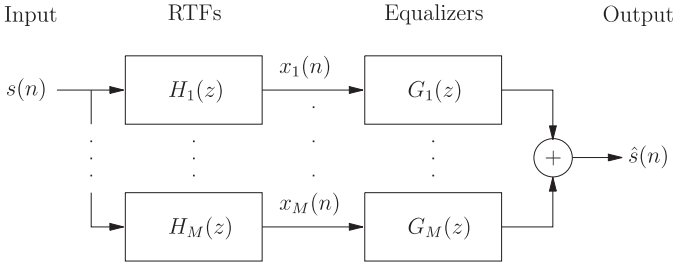


Fig. 1. Full-band multichannel equalization system.

The remainder of the paper is organized as follows. Multichannel equalization is described in Section II. The effects on equalization filter design from inexact RTFs are demonstrated in Section III. The subband equalization method is developed in Section IV. Section V presents a computational complexity analysis of the subband method. Simulation results demonstrating the operation of the proposed algorithm are given in Section VI and, finally, conclusions are drawn in Section VII.

II. Multichannel equalization

The relation in (5) can be written in the time domain as

$$\sum_{m=1}^{M} \mathbf{H}_m \mathbf{g}_m = \mathbf{H}\mathbf{g} = \mathbf{d}, \tag{6}$$

where $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2 \ \dots \ \mathbf{H}_M]$, and $\mathbf{g} = [\mathbf{g}_1^T \ \mathbf{g}_2^T \ \dots \ \mathbf{g}_M^T]^T$. An optimization problem can then be formulated as

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \|\mathbf{H}\mathbf{g} - \mathbf{d}\|^2, \tag{7}$$

and the multichannel equalization (MCEQ) filters can be calculated according to [14]

$$\hat{\mathbf{g}} = \mathbf{H}^{+}\mathbf{d}, \tag{8}$$

where $\mathbf{H}^{+}$ is the matrix pseudo-inverse [22]. The choice of equalization filter length, L_i and, consequently, the dimensions of $\mathbf{H}$, $(L+L_\mathrm{i}-1)\times ML_\mathrm{i}$, define the solution obtained with (8). If $L + L_\mathrm{i} - 1 \leq ML_\mathrm{i}$ then

$$L_\mathrm{i} \geq \frac{L-1}{M-1}, \quad M \geq 2 \tag{9}$$

and the system is underdetermined such that several exact solutions exist [23]. Then the pseudo-inverse in (8) is defined as $\mathbf{H}^{+} = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}$ and gives the minimum norm solution to (7). In the special case when the length in (9) results in an equivalence, the matrix $\mathbf{H}$ becomes square and the pseudo-inverse in (8) reduces to a standard matrix inverse. The exact solution is then unique and equivalent to that of MINT [3]. However, as pointed out in [14], it is not always possible to choose such length for $M > 2$, since the relation in (9) may not give an integer result. Instead, a greater length is often chosen [14], [24]. A third case arises when L_i is chosen such

that $(L+L_{\rm i}-1) > ML_{\rm i}$, which results in an overdetermined system of equations and only a least squares solution can be obtained [23]. For this work, we consider the former, minimum norm exact solutions, and set the equalization filter length to

$$L_{\rm i} = \left\lceil \frac{L-1}{M-1} \right\rceil, \quad M \geq 2, \tag{10}$$

where $\lceil a \rceil$ denotes the ceiling operator giving the smallest integer greater than or equal to a. The relation between an input signal $s(n)$, RTFs $H_m(z)$, equalizers $G_m(z)$, and an output signal $\hat{s}(n)$ is depicted in Fig. 1 where $\hat{s}(n) = \kappa s(n-\tau)$ for ideal equalization.

III. Equalization with inexact impulse responses

In this Section we demonstrate the effects of equalization filter design when using inexact $\mathbf{h}_m$, considering both single channel (approximate) equalization with SCLS and multichannel (exact) equalization with MCEQ. We define an inexact system impulse response, $\tilde{\mathbf{h}}_m = [\tilde{h}_{m,0}\ \tilde{h}_{m,1}\ \ldots\ \tilde{h}_{m,L-1}]^T$, as an impulse response with system mismatch $\mathcal{M}_m > -\infty$ dB, with

$$\mathcal{M}_m = 20\log_{10}\left(\frac{\|\mathbf{h}_m - \tilde{\mathbf{h}}_m\|}{\|\mathbf{h}_m\|}\right) \text{ dB}, \tag{11}$$

where $\|\cdot\|$ denotes Euclidean distance. In the remainder of this work we model system mismatch, as in [25], according to

$$\tilde{\mathbf{h}}_m = (\mathbf{I} + \mathcal{E}_m)\mathbf{h}_m, \tag{12}$$

where $\mathcal{E}_m = \text{diag}\{\epsilon_{m,0}\ \epsilon_{m,1}\ \ldots\ \epsilon_{m,L-1}\}$, $\mathbf{I}$ is the identity matrix, and $\epsilon_{m,i}$ is a zero mean Gaussian variable with the variance set to the desired system mismatch, $\mathcal{M}_m = 10\log_{10}(\text{var}(\epsilon_{m,i}))$ dB.

We now study the design of an equalization filter for $\mathbf{h}_m$ using $\tilde{\mathbf{h}}_m$ when $\mathcal{M}_m > -\infty$ dB. Furthermore, we define the equalized system $\hat{\mathbf{d}} = \mathbf{H}\hat{\mathbf{g}}$ with I-point discrete Fourier transform $\hat{D}(f) = |\hat{D}(f)|e^{j\theta(f)}$, where I is set to the nearest integer power of two larger than the length of $\mathbf{d}$. For evaluation purposes we consider the magnitude and the phase separately as follows.

(i) *Magnitude deviation* is defined here as the standard deviation of the equalized magnitude response [8]

$$\sigma = \sqrt{\frac{1}{I}\sum_{f=0}^{I-1}\left(10\log_{10}|\hat{D}(f)| - \bar{D}\right)^2}, \tag{13}$$

with

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

This measure is scaling independent and equal to zero for exact equalization.

(ii) *Linear phase deviation* is defined as the deviation of the unwrapped phase from a linear fit to its values and is defined here as

$$\Delta = \sqrt{\frac{1}{I}\sum_{f=0}^{N-1}\left(\theta(f) - \bar{\theta}(f)\right)^2}, \tag{14}$$

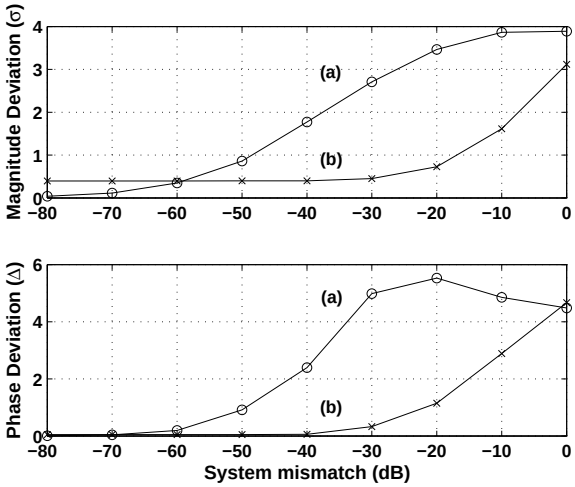


Fig. 2. Magnitude and phase distortion vs. system mismatch for (a) exact equalization with MCEQ from (8) and (b) approximate equalization with SCLS from (4).

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

Two key effects regarding equalization filter design from inexact impulse responses are to be demonstrated: A. the performance degradation caused by increased system mismatch and B. the performance degradation caused by increased system length L for a fixed system mismatch.

A. Effects of system mismatch

An illustrative comparison experiment was performed using an arbitrary system with two random channels $\mathbf{h}_m,\ m = 1, 2$ of length $L = 64$. System mismatch ranging from 0 to -80 dB was modelled using (12). For each case, the impulse response was equalized using the MCEQ method with $L_{\rm i} = L - 1$, $\tau = 0$ and with the SCLS method with $L_{\rm i} = 15L$, $\tau = L/2$. The results, averaged over 100 different random channels, are displayed in Fig. 2. It is seen that equalization using the MCEQ method introduces large spectral distortion for $\mathcal{M}_m > -60$ dB, a level of system mismatch which is the operating range of many current (blind or non-blind) RTF estimation techniques. In contrast, the single channel SCLS equalizer degrades much more gracefully, although equalization filters of very high orders are required. The better performance of the SCLS is a result of the least squares approximation not being able to equalize deep spectral nulls. Furthermore, it is observed that for $\mathcal{M}_m < -70$ dB the multichannel method results in exact equalization while the single channel counterpart reaches a performance bound. These observations are also in accordance with the results reported in [8], [26], where the authors studied equalization of RTF measured at a different location to that at the point of processing.

B. Effects of system length

We examine next the interrelation between system mismatch, impulse response length and equalization accuracy. We consider an arbitrary system with two random channels

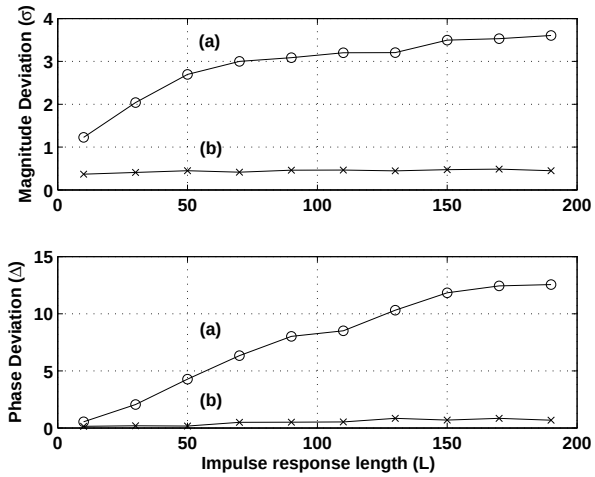


Fig. 3. Magnitude and phase distortion vs. impulse response length for (a) exact equalization with MCEQ from (8) and (b) approximate equalization with SCLS from (4), both with system mismatch $\mathcal{M}_m = -30$ dB.

$\mathbf{h}_m,\ m = 1, 2$ with length L varied in the range 10 to 190 taps and system mismatch $\mathcal{M}_m = -30$ dB. The lengths of the inverse filters were set to $L_\text{i} = L - 1$ and $L_\text{i} = 15L$ for the MCEQ and SCLS equalizers respectively. Figure 3 shows the resulting magnitude and phase distortion for the different channel lengths as an average of 100 different random channel realizations. It can be seen that the exact equalization with MCEQ considerably decreases in performance compared with the single channel SCLS, which appears more or less constant. One reason for the performance degradation with increased system length is that, although the misalignment is kept constant, the total energy of the error in the estimates increases with system length. Moreover, increasing the order of the system results in a larger number of spectral zeros to be equalized, which affects the multichannel equalization in particular since it is more sensitive to errors in the channel estimates compared to the SCLS as seen in Section III-A.

In summary, we have seen that exact multichannel equalization with inverse filters obtained from inexactly estimated systems gives worse results than approximate single channel equalization. However, SCLS inverse filter length of the order $15L$ is not suitable for realistic applications involving acoustic impulse responses and the achieved equalization is limited even when the system mismatch is low. In addition, the deteriorating effects of exact multichannel equalization, for a fixed system mismatch, were seen to increase with increased channel length. These observations lead us to the conclusion that when equalization filters are designed from inexact system estimates, approximate solutions and short system lengths are preferable. The system length due to RTFs is a function of the room and its reverberation time and, therefore, not a controllable system parameter. This motivates the development of a multichannel subband equalizer, where shortened channel length is an inherent feature.

IV. Multichannel subband equalization

We now derive the subband multichannel least squares (SB-MCEQ) equalizer. Figure 4 shows a conceptual system diagram of the SB-MCEQ where the full-band system depicted in Fig. 1 is applied to each subband. This emphasizes three key issues to consider in such design: (i) the choice of the filter-bank, (ii) the mapping of full-band to subband RTFs and (iii) the equalizer design using the subband equivalent filters. Each of these is discussed in the remainder of this Section.

Multirate processing [27] has been applied successfully in acoustic signal processing problems such as, for example, acoustic echo cancellation where significant improvements have been demonstrated in the convergence of the subband adaptive filters [28]–[31]. A subband version of MINT was first investigated in [32]. This approach uses a critically decimated filter-bank. The subband transfer functions to be equalized are estimated using a least squares estimate using the observation of a known reference signal. A different multichannel subband method was proposed by Wang and Itakura [33] for a critically decimated filter-bank. Single channel least squares equalizer is applied to each subband and each microphone and the full-band signal is reconstructed using the best microphone in each subband. The best microphone is selected for each subband using a normalized estimation error criterion from the estimation of the SCLS filters. In [19], a rigourous approach was taken and the relation between full-band and subband filters was studied for an AR model of the room impulse response. An adaptive method for multichannel equalization in oversampled subbands was proposed in [30] and was shown to provide significant improvement over the full-band counterpart.

The relation between full-band and subband filtering was studied, for example, by Lanciani et al [34] for filtering of MPEG audio signals and by Reilly et al [31] with applications to acoustic echo cancellation. The former authors derive the relations between the full-band and subband filters for critically decimated cosine modulated filter banks [27], which are shown to require cross-band filtering. On the other hand, Reilly et al [31] show that good approximations can be obtained with a diagonal filtering matrix, involving only one filter per subband for complex oversampled filter-banks since these sufficiently suppress aliasing in adjacent subbands [30]. We now extend this approach to the multichannel case with application to RTF equalization. This method differs from the previously proposed methods in that it uses oversampled subbands in conjunction with the explicit relation between the full-band and the subband RTFs.

A. Oversampled Filter-banks

The generalized discrete Fourier transform (GDFT) filter-bank [29] is employed in the subsequent development work. The advantages of this filter-bank include straightforward implementation of fractional oversampling and computationally efficient implementations [29]. Within the framework of the GDFT filter-bank, the analysis filters, $u_{k,i}$, are calculated from a single prototype filter, p_i, with bandwidth $\frac{2\pi}{K}$ according to

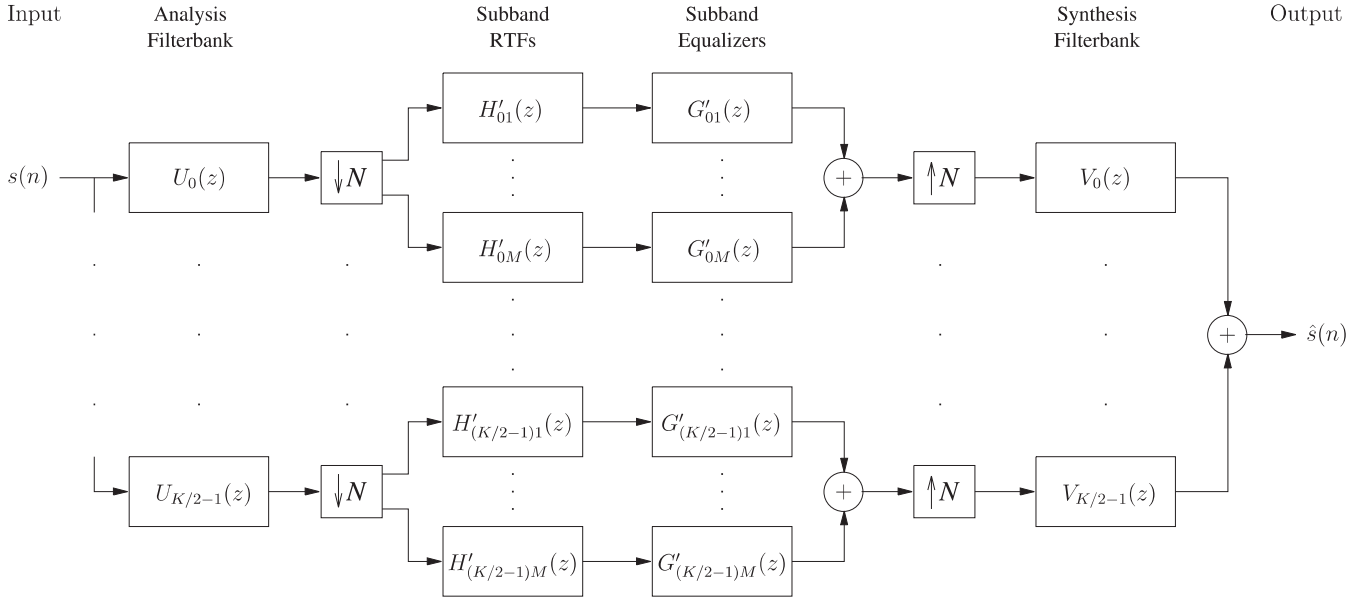


Fig. 4. Subband multichannel equalization system.

the relation [29]

$$u_{k,i} = p_i e^{j\frac{2\pi}{K}(k+k_0)(i+i_0)}, \tag{15}$$

where the properties of the frequency and time offset terms, k_0 and i_0, are discussed in, for example, [29]. We set these to $i_0 = 0$ and $k_0 = 1/2$ as in [31]. It has been shown [29] that a corresponding set of synthesis filters satisfying near perfect reconstruction can be obtained from the time-reversed, conjugated version of the analysis filters

$$v_{k,i} = u^*_{k,L_{\rm pr}-i-1}, \tag{16}$$

where $L_{\rm pr}$ is the length of the prototype filter and, consequently, the length of all analysis and synthesis filters of the filter-bank. Although this filter design results in complex subband signals, for K even, only $K/2$ subbands need to be processed since the remaining subbands are straightforward complex conjugates of these.

The choices of decimation factor, N, and number of subbands, K, has several consequences on the algorithm. A large number of subbands requires a long prototype filter to suppress aliasing effectively. On the other hand, if too few subbands are used, the benefit of shorter subband equalization filters is reduced. The choice of oversampling ratio (N/K) affects the performance of the equivalent subband filters. A good trade off between these parameters was found in the filter-bank used for the illustrative experiments in this paper with $K = 32$ subbands and decimation factor, $N = 24$. An $L_{\rm pr} = 512$-tap prototype filter was designed using the iterative least squares method [29], giving an estimated aliasing suppression of 82 dB. From the properties of the GDFT outlined here, the following two properties can be assumed to be valid:

P1: Aliasing is sufficiently suppressed in the subbands

$$U_k(zW_N^i)V_k(z) \approx 0, \quad i > 0, \ \forall k, \tag{17}$$

where $W_N = e^{-j2\pi/N}$.

P2: Magnitude distortion of the filter-bank is negligible

$$\sum_{k=0}^{K/2-1} U_k(z)V_k(z) \approx \kappa z^{-\tau}, \tag{18}$$

where $U_k(z)$ and $V_k(z)$ are the z-transforms of the subband analysis and synthesis filters respectively.

B. Subband Decomposition

Consider the K subband, M microphone system in Fig. 4. It is clear that, in order to design the subband equalizers $G'_{km}(z)$, the subband RTFs $H'_{km}(z)$ must be found using, for example, complex subband decomposition [31] of their full-band counter parts $H_m(z)$. The objective of the subband decomposition is to find a set of subband filters, $H'_{km}(z),\ k = 0, 1, \ldots, K/2-1$, given the full-band filter $H_m(z)$, such that the total transfer function of the filter bank, $F_m(z)$, is equivalent to the that of the full-band filter up to an arbitrary scale factor, κ, and an arbitrary delay, τ. This can be written

$$F_m(z) = \kappa z^{-\tau} H_m(z), \quad \forall m. \tag{19}$$

The total transfer function of the filter-bank for the mth channel is given by

$$F_m(z) = \frac{1}{N} \sum_{k=0}^{K/2-1} \sum_{i=0}^{N-1} U_k(zW_N^i) H'_{km}(z^N) V_k(z), \tag{20}$$

Evoking property P1 in (17), the filter-bank transfer function reduces to

$$F_m(z) \approx \frac{1}{N} \sum_{k=0}^{K/2-1} U_k(z) H'_{km}(z^N) V_k(z), \tag{21}$$

which allows for a single filter per subband.

Next, following the approach in [31], we choose the filters in each subband, $H'_{km}(z)$, such that they satisfy the relation

$$U_k(z)H'_{km}(z^N) = U_k(z)H_m(z), \quad \forall k. \tag{22}$$

Substituting, (22) into (21) we obtain

$$F_m(z) \approx H_m(z)\frac{1}{N}\sum_{k=0}^{K/2-1} U_k(z)V_k(z). \tag{23}$$

Finally, due to property P2 in (18), we find that the overall filter-bank transfer function is

$$F_m(z) \approx \frac{\kappa}{N} z^{-\tau} H_m(z), \tag{24}$$

which is the desired result. Thus, the remaining problem is to solve for $H'_{km}(z)$ in (22).

Decimating (22) by a factor of N, the following approximation can be formed

$$\frac{1}{N}\sum_{i=0}^{N-1} U_k(z^{1/N}W_N^i)H'_{km}(z) \quad \approx$$
$$\frac{1}{N}\sum_{i=0}^{N-1} U_k(z^{1/N}W_N^i)H_m(z^{1/N}W_N^i), \tag{25}$$

which in the time domain is equivalently written

$$\mathbf{U}_{N,k}\mathbf{h}'_{km} \approx \mathbf{r}_{N,km} \tag{26}$$

where $\mathbf{h}'_{km} = [h'_{km,0} \;\; h'_{km,1} \;\; \dots \;\; h'_{km,L'-1}]^T$ are the L'-tap subband impulse responses, $\mathbf{r}_{N,km} = [r_{km,0} \; r_{km,N} \; \dots \; r_{km,N(L-1)}]^T$ is an $\lceil (L+L_{\rm pr}-1)/N \rceil \times 1$ vector with $r_{km,i} = h_{m,i} * u_{k,i}$ and

$$\mathbf{U}_{N,k} = \begin{bmatrix} u_{k,0} & 0 & \dots & 0 \\ u_{k,N} & u_{k,0} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ u_{k,L_{\rm pr}-1} & \dots & \vdots & 0 \\ 0 & u_{k,L_{\rm pr}-1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & u_{k,L_{\rm pr}-1} \end{bmatrix},$$

where $L_{\rm pr}$ is the length of the analysis filters. The convolution on the left hand side of (26) is of length $\lceil L_{\rm pr}/N \rceil + L' - 1$, and consequently, the length of the subband filters is

$$L' = \left\lceil \frac{L+L_{\rm pr}-1}{N} \right\rceil - \left\lceil \frac{L_{\rm pr}}{N} \right\rceil + 1 \tag{27}$$

The estimates of the subband filters $\hat{\mathbf{h}}'_{km}$ are then found by solving the following optimization problem [31]

$$\hat{\mathbf{h}}'_{km} = \arg\min_{\mathbf{h}'_{km}} \|\mathbf{U}_{N,k}\mathbf{h}'_{km} - \mathbf{r}_{N,km}\|^2. \tag{28}$$

The kth subband, mth channel filters are then obtained in the least squares optimal sense according to

$$\hat{\mathbf{h}}'_{km} = \mathbf{U}^+_{N,k}\mathbf{r}_{N,km}. \tag{29}$$

In summary, given a full-band RTF, $H_m(z)$, and $K/2$-band filter bank satisfying perfect reconstruction and aliasing suppression in the subbands, a set of subband filters, $H'_{km}(z)$, of order L/N, can be found such that the overall subband transfer function is equivalent to the full-band filter response. We now aim to exploit the significant order reduction in the subbands of the very long full-band room impulse responses.

C. Subband multichannel equalization

The multichannel equalization filters, $\hat{G}'_{km}(z)$, can be calculated for each subband using the filters $\hat{H}'_{km}(z)$ obtained from (29). Here, this is done utilizing the multichannel equalization filter design from (8), which now becomes

$$\hat{\mathbf{g}}'_k = \hat{\mathbf{H}}'^{+}_k \mathbf{d}, \quad k = 0, 1, \dots, \frac{K}{2} - 1, \tag{30}$$

such that for each subband

$$\sum_{m=1}^{M} \hat{G}'_{km}(z)\hat{H}'_{km}(z) = 0, \quad k = 1, 2, \dots, \frac{K}{2} - 1. \tag{31}$$

Thus, equalization is achieved by applying the inverse filters, $\hat{\mathbf{g}}'_k$, to the subband signals of the reverberant observations in each subband k, $\forall k$ and an equalized full-band signal is constructed. Assuming that exact equalization is achieved in each subband, the accuracy of the final result will depend on the reconstruction properties of the filter-bank, the level of aliasing suppression and, consequently, on the design of the prototype filter. Therefore, the overall equalization of the subband method will not be exact in practice, which can be beneficial as discussed in Section III. These dependencies will be explained through illustrative simulations in Section VI.

V. Computational complexity

In this Section we present a comparative analysis of the number of computations required for the solution of the full-band MCEQ equalizer design and the SB-MCEQ equalizer design (including the computational cost of the subband decomposition). The comparison is made in terms of floating point operations (flops), where one flop is defined as either one real multiplication or one real addition [22]. We consider the general optimization problem $\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2$, which has a minimum norm solution $\hat{\mathbf{x}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$, 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 using the normal equations is given by [22]

$$pq^2 + \frac{q^3}{3}. \tag{32}$$

From the dimensions of the full-band equalization filter calculation in (8), the number of flops required for the MCEQ design is

$$(ML_{\rm i})^2(L + L_{\rm i} - 1) + \frac{(ML_{\rm i})^3}{3}. \tag{33}$$

The subband equalization filter design takes into consideration two separate calculations for each of the $K/2$ subbands: the cost of the subband inverse filter computation in (30) and the cost of the subband decomposition in (29). 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. Under

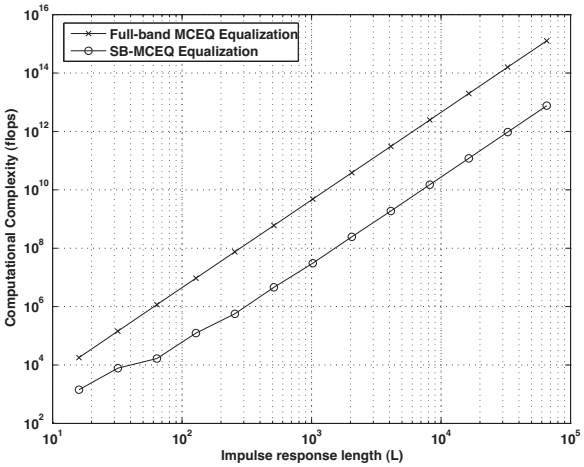


Fig. 5. Floating point operation count vs. system length for the full-band ($\times$) and subband ($\circ$) equalizers.

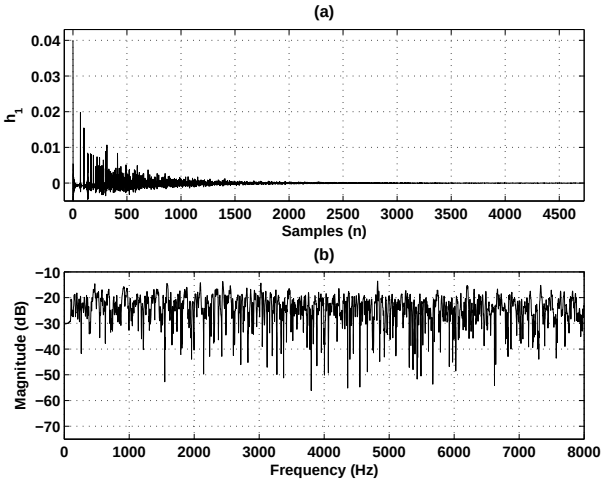


Fig. 6. Typical example of (a) a simulated room impulse response and (b) the corresponding magnitude response.

the assumption that an equal number of complex multiplications and complex additions are required to solve the system of equations considered here, we multiply the expression in (32) by a factor of four. The total flops required for the subband inverse filter design can be expressed as

$$2K\left((ML_{\mathrm{i}}')^2(L'+L_{\mathrm{i}}'-1)+(L_rL')^2+\frac{(ML_{\mathrm{i}}')^3+L'^3}{3}\right). \tag{34}$$

The key factor of the computational complexity is the system length and thus, the improvement achieved by the subband method will depend on the number of subbands and on the decimation ratio. An example is given in Fig. 5 where the computational complexity is calculated with (33) and (34) respectively. The subband implementation for this example is that presented in Section IV-B with $K = 32$ subbands decimated by $N = 24$. On average over all lengths, the subband approach reduces the computational complexity by a factor of 120.

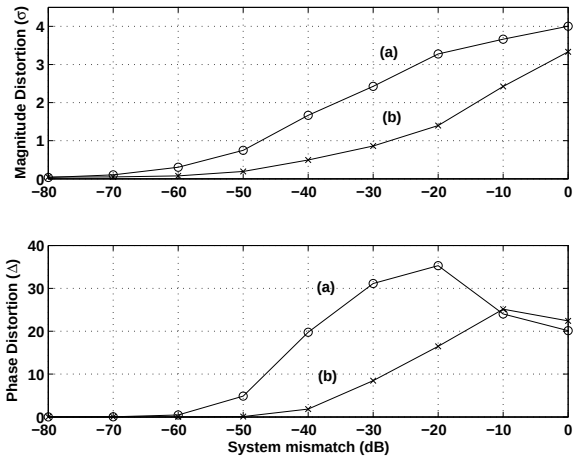


Fig. 7. Magnitude and phase distortions vs. system mismatch for SB-MCEQ equalization of simulated room impulse responses.

VI. Simulations and Results

The following simulation results are presented to demonstrate the performance of the proposed SB-MCEQ equalization method. Three experiments were performed to show: (i) a comparative performance evaluation with the full-band MCEQ using randomly generated channels, (ii) the applications of the SB-MCEQ to simulated room impulse responses and (iii) illustrative examples of the algorithm.

A. Experiment 1: Simulated RTFs

The experiment demonstrates the performance of the SB-MCEQ equalizer, compared with the full-band MCEQ using simulated RTFs. A linear array of $M = 5$ uniformly distributed microphones with 0.1 m separation between adjacent sensors was simulated using the source-image method [11] for a room with dimensions $6\times5\times3$ m. The impulse response at one of the microphones, $\mathbf{h}_1$, and the corresponding magnitude response are depicted in Figs. 6(a) and 6(b) respectively. The sampling frequency was $f_{\mathrm{s}} = 16$ kHz and the room reverberation time was $T_{60} = 0.3$ s, resulting channel lengths of $L = 4800$ taps. Moreover, keeping the source-microphone configuration fixed, RTFs were simulated at 100 different locations in the room. System misalignment, $\mathcal{M}_m$, varying between 0 and -80 dB was simulated with (12). The full-band equalization filters in (8) were computed with the SLICOT toolbox [35] according to the method discussed in [23].

Figure 7 shows the results in terms of magnitude and phase distortion, as an average of the 100 measurement locations for (a) the full-band MCEQ and (b) for the proposed subband implementation. Notably, the SB-MCEQ exhibits much more graceful performance degradation with increased misalignment compared to the full-band MCEQ and with a similar behaviour as the single channel SCLS equalizer shown

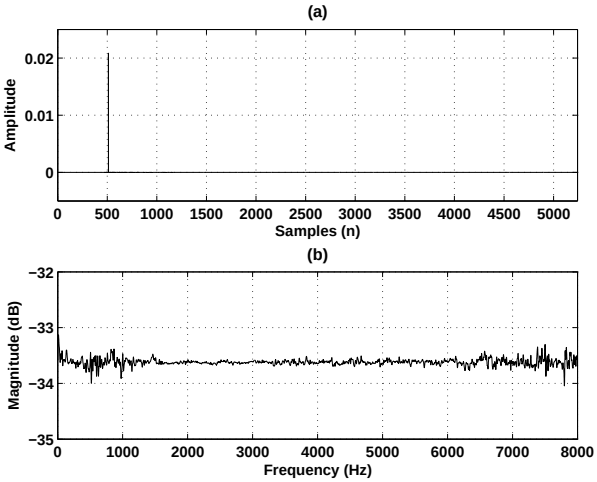


Fig. 8. Equalized (a) time domain impulse response and (b) magnitude response, using the SB-MCEQ method $\mathcal{M}_m = -80$ dB. The magnitude distortion is $\sigma = 0.03$. (Note that the magnitude scaling of the equalized impulse response is of no significance.)

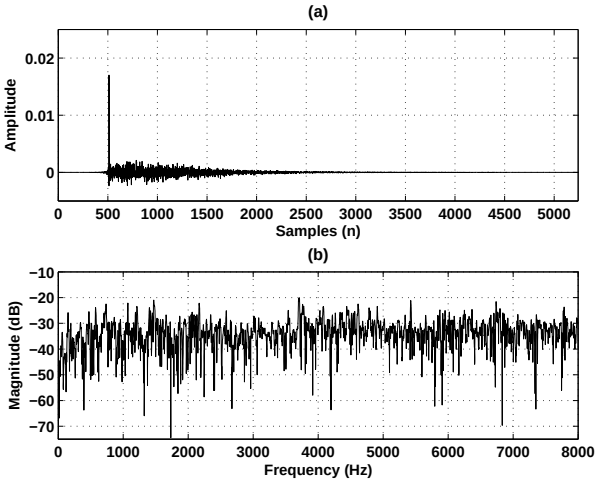


Fig. 9. Equalized (a) time domain impulse response and (b) magnitude response, using the SB-MCEQ method for $\mathcal{M}_m = -10$ dB. The magnitude distortion is $\sigma = 2.63$. (Note that the magnitude scaling of the equalized impulse response is of no significance but the relative scaling between Figs. 8a and 9a is significant.)

in Fig. 2. Thus, the SB-MCEQ method is shown in these results to be less sensitive to inexact impulse responses, while benefiting from the shorter filters of multichannel inversion. This improvement is a consequence of the reduced filter length in the subbands, which in Section III-B was demonstrated to improve the MCEQ equalizer performance. In addition, nearly perfect equalization is achieved with the SB-MCEQ method for $\mathcal{M}_m \leq -40$ dB.

Finally, we provide two characteristic examples of the subband equalizer output for the simulated RTFs. Figure 8a shows a typical outcome of the equalized room impulse response in the time domain and Fig. 8b shows the corresponding magnitude response for $\mathcal{M}_m = -80$ dB. It can be seen that near perfect equalization is achieved with only small spectral distortion ($\sigma = 0.03$); this distortion results from the approximations in the subband filter decomposition and in the filter-bank reconstruction. Thus, the accuracy depends on the ability of the prototype filter to suppress aliasing and on the oversampling ratio. The delay in the equalized impulse in Fig. 8a is due to the filter-bank and is governed by the order of the prototype filter $L_{\rm pr}$. As a further illustration for a less accurate RTF estimation, a characteristic outcome for $\mathcal{M}_m = -10$ dB is shown in Fig. 9, where a more significant spectral distortion is observed, which is due to the room impulse response inaccuracies.

B. Experiment 2: Speech Dereverberation

In Experiment 2, we used the impulse responses and the equalizing filters from Experiment 1 and applied these to speech dereverberation. The sentence 'Hoist the load to your left shoulder.' uttered by a male talker, drawn from the IEEE corpus [36], was used as an example. The Segmental Signal to Reverberation Ratio (SRR) [1] was used as an objective evaluation metric. The results, averaged over 100 different source-microphone configurations, are shown in Fig. 10 for

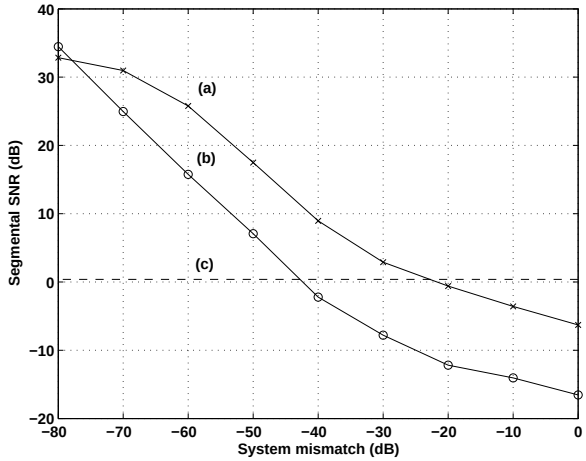


Fig. 10. Segmental SRR for (a) speech equalized with SB-MCEQ, (b) speech equalized with full-band MCEQ and (c) unprocessed reverberant speech at one channel.

(a) speech equalized with the proposed subband approach, (b) speech equalized with the full-band MCEQ and (c) unprocessed speech at the microphone closest to the talker. It can be seen that equalizing with channel estimates with misalignment larger than $\mathcal{M}_m - 40$ dB results in lower Segmental SRR than in that of the unprocessed reverberant signal. The reduced sensitivity to errors in the channel estimates of the subband SB-MCEQ method is manifested here by showing that equalization can be beneficial down to misalignments of $\mathcal{M}_m - 20$ dB; for this example, there is, on average over all misalignments, 9 dB improvement in Segmental SRR using the subband method compared to the full-band method.

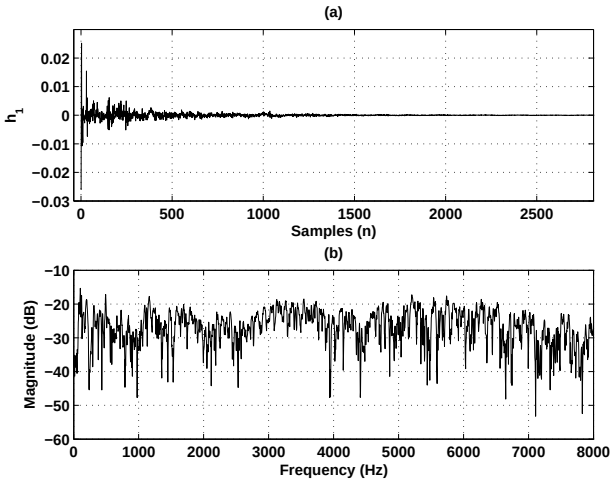


Fig. 11. A measured (a) room impulse response and (b) the corresponding magnitude response.

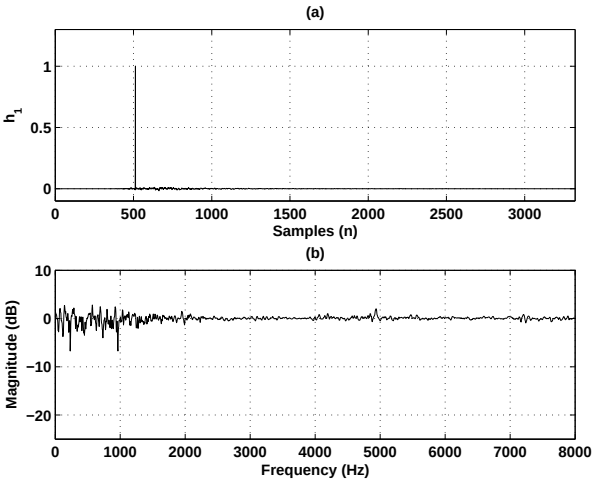


Fig. 13. Equalized (a) time domain impulse response and (b) magnitude response, using the SB-MCEQ method for $\mathcal{M}_m = -50$ dB.

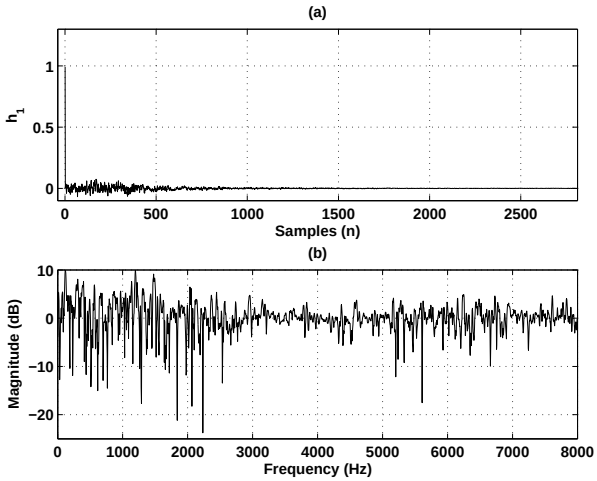


Fig. 12. Equalized (a) time domain impulse response and (b) magnitude response, using the full-band MCEQ method for $\mathcal{M}_m = -50$ dB.

C. Experiment 3: Measured RTFs

Finally, we provide an example of equalization using measured RTFs obtained from the MARDY database [37]. An example (a) impulse response and (b) the corresponding magnitude response are shown in Fig. 11. System misalignment corresponding to $\mathcal{M}_m = -50$ dB was simulated with (12) and the resulting RTFs were employed in the design the equalization filters using both the full-band and the subband methods. The equalized RTF using the full-band MCEQ and the SB-MCEQ are shown in Fig. 12 and Fig. 13, respectively. The smaller spectral distortion caused by the subband method is conspicuous.

VII. Conclusions

Equalization of acoustic impulse responses has been discussed both for single and multiple microphones. Single microphone approaches can provide only approximate equalization, require very long inverse filters and result in long processing delay due to the non-minimum phase property of the RTFs. On the other hand, exact equalization with no delay and with inverse filters of similar order to the room impulse responses is possible in the multimicrophone case. However, multichannel methods are very sensitive to inaccuracies in the estimated systems to be equalized, causing significant distortions to the equalized signal.

Consequently, a new algorithm was derived operating on decimated oversampled subband signals, where the full-band impulse response is decomposed into equivalent filters in the subbands and multichannel least squares equalization is applied to each subband. It was shown that this method results in substantial computational savings at the cost of very small spectral distortion due to the filter bank. Simulation results were presented to evaluate the performance of this method and equalization of channels of several thousand taps was demonstrated. Most importantly, experimental results indicated that the new method is more robust to errors in the impulse responses of the system to be equalized, which is due to a combination of shorter filters and approximation of the filtering in the subbands. Thus, the proposed subband multichannel least squares equalization benefits from the reduced sensitivity to channel estimation errors, shorter equalization filters, no delay due to the equalization (the delay due to the filter bank is less than 32 ms in our examples), giving significant advantages over existing single and multichannel techniques.

References

[1] P. A. Naylor and N. D. Gaubitch, "Speech Dereverberation," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Eindhoven, The Netherlands, Sept. 2005.

[2] P. A. Nelson, F. Orduña-Brustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 3, pp. 185–192, Nov. 1995.

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

[4] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, Prentice Hall, 3 edition, 1996.

[5] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Amer.*, vol. 66, no. 1, pp. 165–169, July 1979.

[6] H. Kuttruff, *Room Acoustics*, Taylor & Francis, 4 edition, Oct. 2000.

[7] M . R. Schroeder, "Statistical Parameters of the Frequency Response Curves of Large Rooms," *J. Audio Eng. Soc.*, vol. 35, no. 5, pp. 299–305, May 1987.

[8] 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.

[9] J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 1982, vol. 7, pp. 1858–1861.

[10] J. N. Mourjopoulos, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42, no. 11, pp. 884–900, Nov. 1994.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[12] B. D. Radlović and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 6, pp. 728–737, Nov. 2000.

[13] M. Tohyama, R. H. Lyon, and T. Koike, "Source waveform recovery in a reverberant space by cepstrum dereverberation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Apr. 1993, vol. 1, pp. 157–160.

[14] 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.

[15] S. Bharitkar, P. Hilmes, and C. Kyriakakis, "Robustness of spatial average equalization: A statistical reverberation model approach," *J. Acoust. Soc. Amer.*, vol. 116, no. 6, pp. 3491–3497, Dec. 2004.

[16] M. Hofbauer and H. Loeliger, "Limitations for FIR multi-microphone speech dereverberation in the low-delay case," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sept. 2003, pp. 103–106.

[17] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 2, pp. 320–328, Apr 1994.

[18] Y. Haneda, S. Makino, and Y. Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 4, pp. 325–333, Jul 1997.

[19] J. R. Hopgood and P. J. W. Rayner, "A probabilistic framework for subband autoregressive models applied to room acoustics," in *Proc. IEEE Workshop Statistical Signal Processing*, Aug 2001, pp. 492 – 495.

[20] T. Hikichi, M. Delcroix, and M. Miyoshi, "On Robust Inverse Filter Design for Room Transfer Function Fluctuations," in *Proc. European Signal Processing Conf.*, Sept. 2006.

[21] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sept. 2006, pp. 1–4.

[22] G. H. Golub and C. F. van Loan, *Matrix computations*, John Hopkins series in the mathematical sciences. London: John Hopkins University Press, 3 edition, 1996.

[23] M. Hofbauer, *Optimal Linear Separation and Deconvolution of Acoustical Convolutive Mixtures*, Ph.D. thesis, The Swiss Federal Institute of Technology (ETH), Zürich, 2005.

[24] 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, pp. 1–12, 2007.

[25] 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.

[26] F. Talantzis and D. B. Ward, "Robustness of multi-channel equalization in an acoustic reverberant environment," *J. Acoust. Soc. Amer.*, vol. 114, no. 2, pp. 833–841, Aug. 2003.

[27] P. P. Vaidyanathan, *Multirate systems and filter banks*, Prentice Hall, 1993.

[28] P. A. Naylor, O. Tanrikulu, and A. G. Constantinides, "Subband adaptive filtering for acoustic echo control using allpass polyphase IIR filterbanks," *IEEE Trans. Speech Audio Processing*, vol. 6, no. 2, pp. 143–155, Mar. 1998.

[29] S. Weiss and R. W. Stewart, *On adaptive filtering in oversampled subbands*, Shaker Verlag, 1998.

[30] S. Weiss, G. W. Rice, and R. W. Stewart, "Multichannel Equalization in Subbands," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1999, pp. 203–206.

[31] 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.

[32] H. Yamada, H. Wang, and F. Itakura, "Recovering of broad band reverberant speech signal by sub-band MINT method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1991, pp. 969–972.

[33] H. Wang and F. Itakura, "Realization of Acoustic Inverse Filtering through Multi-Microphone Sub-Band Processing," *IEICE Trans. Fundamentals*, vol. E75-A, no. 11, pp. 1474–1483, Nov. 1992.

[34] C. A. Lanciani and R. W. Schafer, "Subband-domain filtering of MPEG audio signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Mar. 1999, vol. 2, pp. 917–920.

[35] A. Varga and P. Benner, "SLICOT - a subroutine library in systems and control theory," in *Applied and Computational Control, Signal and Circuits*, B. N. Datta, Ed., vol. 1, pp. 499–539. Birkhäuser, 1999.

[36] IEEE Subcommittee, "IEEE Recommended Practice for Speech Quality Measurements," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, no. 3, pp. 225–246, 1969.

[37] 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*, Paris, France, Sept. 2006.