OPTIMAL BEAMFORMING AS A TIME DOMAIN EQUALIZATION PROBLEM WITH APPLICATION TO ROOM ACOUSTICS

Mark R. P. Thomas, Ivan J. Tashev

Microsoft Research
Redmond, WA 98052, USA
{markth, ivantash}@microsoft.com

Felicia Lim, Patrick A. Naylor

Dept. of Electrical and Electronic Engineering
Imperial College London, UK
{felicia.lim06, p.naylor}@imperial.ac.uk

ABSTRACT

Signals captured by microphone arrays provide spatial diversity that can be exploited by multichannel processing algorithms to suppress noise and reverberation. Beamforming is a class of approaches that treats the problem with respect to the spatial location of wanted and competing sources, leveraging properties of propagation of waves in free space. A related class of algorithms is channel equalization that exploits knowledge of the acoustic impulse response between a source and microphones with a view to near-perfect dereverberation. Beamforming has been shown to be a very powerful and practical tool in a number of domains, whereas channel equalizers are notoriously sensitive to noise and channel mismatch leading to limited practical applicability. This paper investigates some of the common properties of these algorithms and presents a solution incorporating approaches from both disciplines.

Index Terms— Beamforming, channel equalization, dereverberation

1. INTRODUCTION

Speech signals captured by hands-free devices are typically corrupted by noise and reverberation that impair the perceived quality and intelligibility of the received speech. Microphone arrays are attractive because the signals they receive exhibit spatial diversity that can be exploited to suppress components that are not spatially collocated with the source. Processing of multichannel signals can be divided into two broad classes: beamforming [1], which is largely motivated by the propagation of waves in free space and/or the spatial correlation of noise signals, and channel equalization [2], motivated by the acoustic impulse response between the source and receivers in a reverberant environment.

Beamformer design often operates under certain common assumptions: that all sources lie in the farfield, all sensors are omnidirectional, and propagation from a source to the array is characterized by a pure delay. This is particularly true in the design of non-adaptive superdirective beamformers [3]. Some assumptions are circumvented with adaptive (data-dependent) solutions, however both adaptive and non-adaptive approaches often incorporate distortionless constraints to ensure that waves propagating in the wanted direction are left undistorted under the assumption that the source-receiver transfer function is a pure delay. Conversely, channel equalizers such as the Multichannel Input-Output Inverse Theorem (MINT) [2] use prior knowledge of the reverberant impulse responses from source to receivers with a view to near-perfect equalization, but without explicit constraints for spatial selectivity. Channel inversion techniques are notoriously sensitive to errors in channel estimates and can often increase the level of reverberation under channel mismatch, for example due to a different source location, the relocation of furniture, or a change in the ambient temperature [4, 5]. To this end, channel shortening/reshaping techniques [6, 7] have been proposed to improve robustness.

Both beamforming and channel inversion can be viewed as filter-and-sum operations with different optimization criteria. The concept of MINTForming [8] was introduced to combine both concepts into a single algorithm that controls the tradeoff between the channel inversion provided by the MINT algorithm and the spatial and noise performance of an optimal filter-and-sum beamformer. The approach was to formulate both MINT and beamforming as frequency domain design problems, to combine their respective cost functions, and to evaluate the performance as a channel equalizer.

In order to apply frequency domain designs to real world signals it is necessary to obtain a finite impulse response (FIR) approximation [9] to the frequency domain filter. This approximation inherently introduces error into the design, producing suboptimal behavior. Several works have proposed time-domain beamformer designs to circumvent this issue [9, 10, 11]. Conversely, channel equalization algorithms such as MINT are also usually posed as time-domain problems [2, 4, 7] so no approximation is required.

This paper considers MINT and non-adaptive superdirective beamforming purely as both time domain design problems. As a reference, a third hybrid case is considered whereby a beamformer is designed using knowledge of the reverberant impulse response for sources in several locations. The performance is evaluated both in terms of channel equalization in the wanted direction and in terms of spatial selectivity.

The remainder of this paper is organized as follows. The equalization and beamforming problems are formulated in Sec. 2. In Section 3, MINT, an optimal filter-and-sum beamformer and a hybrid reference algorithm are formulated in the time domain. The algorithms are evaluated in Section 4 and conclusions are drawn in Section 5.

2. PROBLEM FORMULATION

Consider an array of M microphones placed in a reverberant environment. Let

$$\check{\mathbf{h}}_{p,m} = [\check{h}_{p,m}(0)\ \ldots\ \check{h}_{p,m}(L-1)]^T \in \mathbb{R}^{L\times 1} \quad (1)$$

$$\vec{\mathbf{h}}_{p,m} = [\vec{h}_{p,m}(0)\ \ldots\ \vec{h}_{p,m}(L-1)]^T \in \mathbb{R}^{L\times 1} \quad (2)$$

be the impulse responses of length L samples between a source with position index $p \in \{1, 2, \ldots, P\}$ and receiver $m \in \{1, 2, \ldots, M\}$

in the reverberant and anechoic cases respectively[1]. The source at index p_0 is considered to be a wanted source and all other values of p are considered to be unwanted noise sources. It is assumed that $\check{\mathbf{h}}_{p,m}$ and $\vec{\mathbf{h}}_{p,m}$ contain all propagation delays and have not been truncated. Additionally, let

$$\check{\mathbf{h}}_p = [(\check{\mathbf{h}}_{p,1})^T \dots (\check{\mathbf{h}}_{p,M})^T]^T \in \mathbb{R}^{ML \times 1} \quad (3)$$

$$\vec{\mathbf{h}}_p = [(\vec{\mathbf{h}}_{p,1})^T \dots (\vec{\mathbf{h}}_{p,M})^T]^T \in \mathbb{R}^{ML \times 1} \quad (4)$$

contain stacked impulse responses between source at p and all M microphones. The aim for both filter-and-sum beamformer and channel equalizer design is to synthesize filters

$$\mathbf{g}_m = [g_m(0) \ \dots \ g_m(L_i - 1)]^T \in \mathbb{R}^{L_i \times 1} \quad (5)$$

that produce a desired response at the system output, which will be referred to in this paper as *equalization* filters irrespective of whether they were designed for beamforming or equalization. Their stacked representation over all microphones is defined in a similar way to (3)

$$\mathbf{g} = [\mathbf{g}_1^T \dots \mathbf{g}_M^T]^T \in \mathbb{R}^{ML_i \times 1}. \quad (6)$$

Now let $\check{\mathbf{H}}_{p,m} \in \mathbb{R}^{(L+L_i-1) \times L_i}$ be a convolution matrix derived from $\check{h}_{p,m}$ so that $\check{\mathbf{H}}_{p,m}\mathbf{g}_m$ and $h_{p,m}(n) * g_m(n)$, where $*$ denotes linear convolution, are equivalent:

$$\check{\mathbf{H}}_{p,m} = \begin{bmatrix} h_{p,m}(0) & 0 & \cdots & 0 \\ h_{p,m}(1) & h_{p,m}(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_{p,m}(L-1) & \cdots & \vdots & \vdots \\ 0 & h_{p,m}(L-1) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & h_{p,m}(L-1) \end{bmatrix}. \quad (7)$$

A similar formulation is used for $\vec{\mathbf{H}}_{p,m}$. Convolution matrices can be stacked for all M channels,

$$\check{\mathbf{H}}_p = [\check{\mathbf{H}}_{p,1} \ \cdots \ \check{\mathbf{H}}_{p,M}] \in \mathbb{R}^{(L+L_i-1) \times ML_i} \quad (8)$$

$$\vec{\mathbf{H}}_p = [\vec{\mathbf{H}}_{p,1} \ \cdots \ \vec{\mathbf{H}}_{p,M}] \in \mathbb{R}^{(L+L_i-1) \times ML_i}, \quad (9)$$

then further stacked over all P source positions to form universal convolution matrices

$$\check{\mathbf{H}} = [\check{\mathbf{H}}_1^T \dots \check{\mathbf{H}}_P^T]^T \in \mathbb{R}^{P(L+L_i-1) \times ML_i} \quad (10)$$

$$\vec{\mathbf{H}} = [\vec{\mathbf{H}}_1^T \dots \vec{\mathbf{H}}_P^T]^T \in \mathbb{R}^{P(L+L_i-1) \times ML_i}. \quad (11)$$

The equalization filters $\mathbf{g}$ are then synthesized in such a way that the response of the equalized system in direction p is found by a filter-and-sum operation:

$$\mathbf{y}_p = \sum_{m=1}^{M} \check{\mathbf{H}}_{p,m}\mathbf{g}_m = \check{\mathbf{H}}_p\mathbf{g} \in \mathbb{R}^{(L+L_i-1) \times 1}, \quad (12)$$

where $\mathbf{y}_p = [y_p(0) \dots y_p(L + L_i - 2)]^T$. For notational convenience, the equalized output can be found for all P directions in a single operation

$$\mathbf{y} = \check{\mathbf{H}}\mathbf{g} \in \mathbb{R}^{P(L+L_i-1) \times 1} \in \mathbb{R}^{P(L+L_i-1) \times 1}, \quad (13)$$

where $\mathbf{y} = [\mathbf{y}_1^T \ \mathbf{y}_2^T \dots \mathbf{y}_P^T]^T$.

3. ALGORITHMS

3.1. MINT

The Multichannel Input-Output Inverse Theorem (MINT) algorithm [2] was proposed as a means of providing exact inverse filtering of room acoustics from multichannel observations. A stable, causal inverse of a single-channel system does not generally exist due to the nonminimum phase characteristic of room transfer functions. In the multichannel case, it was shown that under certain conditions a stable, finite, causal and exact inverse always exists.

Considering only the wanted direction p_0, equalization filters are defined using MINT to produce the equalized output

$$d_{p_0}(l) = \begin{cases} 1 & \text{if } l = \tau; \\ 0 & \text{otherwise,} \end{cases} \quad (14)$$

where τ is an arbitrary integer delay with vector representation

$$\mathbf{d}_{p_0} = [d_{p_0}(0) \cdots d_{p_0}(L + L_i - 2)]^T \in \mathbb{R}^{(L+L_i-1) \times 1}. \quad (15)$$

The equalizer design can then be stated as a least-squares convex optimization problem

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} ||\check{\mathbf{H}}_{p_0}\mathbf{g} - \mathbf{d}_{p_0}||_2^2, \quad (16)$$

yielding theoretically exact equalization providing the filters $\check{\mathbf{h}}_{m,p_0}$ are known, that there are no common zeros between $\check{\mathbf{h}}_{m,p_0}(n)$ and $\check{\mathbf{h}}_{m+1,p_0}(n)$, and

$$L_i \geq \left\lceil \frac{L-1}{M-1} \right\rceil \quad (17)$$

is satisfied [2].

3.2. Optimal Filter-and-Sum Beamformer (FSB)

In a typical superdirective beamforming application, it is desirable to have a response $\mathbf{y}$ yielding unit gain in the look direction p_0 and minimized gain elsewhere. Capture vectors describing the array behavior for a source in look direction p are usually defined in terms of pure delays of the direct-path signal. In the time domain the filter design can be stated as the convex optimization problem with *distortionless* constraint

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} ||\vec{\mathbf{H}}\mathbf{g} - \mathbf{d}||_2^2 \ \text{ subject to } \vec{\mathbf{H}}_{p_0}\mathbf{g} = \mathbf{d}_{p_0}, \quad (18)$$

where the desired spatial desired response $\mathbf{d}$ is

$$\mathbf{d} = [\mathbf{0} \dots \mathbf{d}_{p_0}^T \dots \mathbf{0}]^T \in \mathbb{R}^{P(L+L_i-1) \times 1}. \quad (19)$$

Other desired responses such as a tapering towards the look direction can help to reduce the amplitude of sidelobes. Time domain formulations for filter-and-sum beamformers are not common; it is derived in this way to make this and MINT mutually compatible.

3.3. Oracle Case

Using a similar approach to the optimal beamformer, the problem can be restated assuming knowledge of reverberant impulse responses $\check{h}_p$ in all P directions, thereby using all available information:

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} ||\check{\mathbf{H}}\mathbf{g} - \mathbf{d}||_2^2 \ \text{ subject to } \check{\mathbf{H}}_{p_0}\mathbf{g} = \mathbf{d}_{p_0}. \quad (20)$$

[1] Accents $(\check{\cdot})$ and $(\vec{\cdot})$ depict a reflection and the direct path.

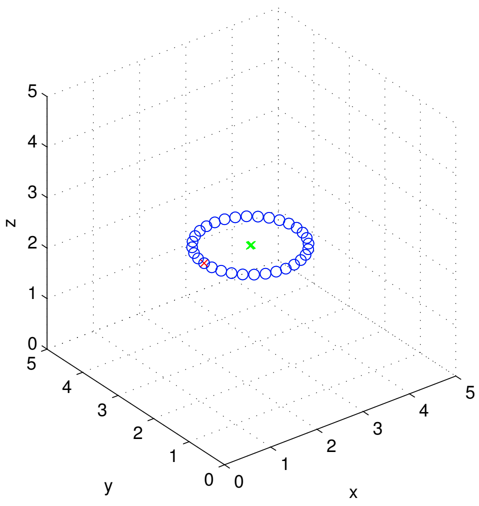

Fig. 1. Microphone array (green ×), evaluation points (blue ∘) and look direction (red ×).

Notice the constraint in (20) is identical to the MINT requirement in (16). In the case that the filter length requirement (17) is just satisfied, it is clear that the constraint in (20) consumes every available degree of freedom so that (20) and (16) will yield the same filter. In order to introduce spatial selectivity, either the equalizer length L_i should be increased to increase the available degrees of freedom, or an inequality introduced to the constraint as

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \|\check{\mathbf{H}}\mathbf{g} - \mathbf{d}\|_2^2 \text{ subject to } \|\check{\mathbf{H}}_{p_0}\mathbf{g} - \mathbf{d}_{p_0}\|_2^2 < \epsilon, \quad (21)$$

where ϵ is an arbitrary constant.

4. EVALUATION

The aim of this experiment is to evaluate the algorithms under test in ideal conditions, i.e. when the impulse responses are known exactly and no robustness constraints are applied. This was chosen to provide insight into shortcomings that limit their ultimate performance. Defining all three cases as time domain problems also helps to makes informed comparisons.

4.1. Metrics

White noise gain measures the sensitivity of the approaches to sensor noise. Assuming the distortionless constraint is met,

$$\text{WNG} = -10\log_{10}\left(\mathcal{g}(k)^H \mathcal{g}(k)\right), \quad (22)$$

where $\mathcal{g}(k) = [\mathcal{g}_1(k)\dots\mathcal{g}_M(k)]^T \in \mathbb{C}^{M\times 1}$ is a vector of discrete Fourier transforms of $g_m(n)$ and $(\cdot)^H$ is a Hermitian (conjugate) transpose. The algorithms under test should be evaluated both as channel equalizers and a beamformers since they draw upon ideas from both fields. To this end, the equalized impulse response (EIR) is first calculated for all look directions under reverberant conditions:

$$\mathbf{y} = \check{\mathbf{H}}\hat{\mathbf{g}}. \quad (23)$$

Ideally $\|\mathbf{y}_{p_0} - \mathbf{d}_{p_0}\|_2^2 = 0$ and $\|\mathbf{y}_{p\neq p_0}\|_2^2 \simeq 0$. There are several measures derived from (23) for a single source [12]. Here we shall

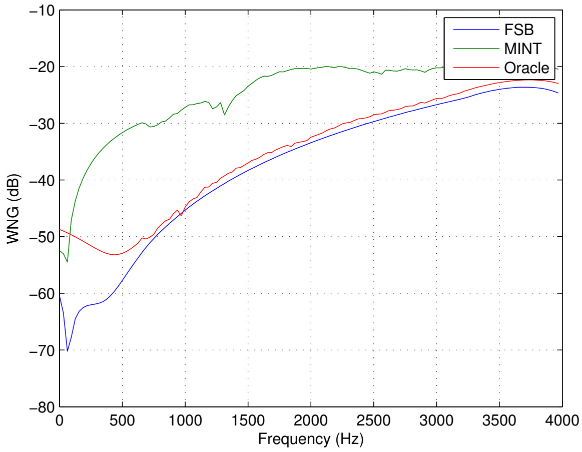

Fig. 2. White noise gains.

use the *Direct to Reverberant Ratio* (DRR)

$$\text{DRR}_p = 10\log_{10}\left(\frac{y_p^2(\tau)}{\sum_{n=0}^{\tau-1} y_p^2(n) + \sum_{n=\tau+1}^{L+L_i-2} y_p^2(n)}\right) \text{dB}, \quad (24)$$

in which the direct component is defined as a single sample at index τ. As a measure of beamformer performance, spatial selectivity is often evaluated with the *directivity index* (DI) that measures the ratio of the sensitivity in the look direction to the mean of sensitivity over the entire space. Letting $\mathcal{y}_p(k)$ be the discrete Fourier transform of $y_p(n)$, the *Reverberant Directivity Index* (RDI) is

$$\text{RDI}(k) = 10\log_{10}\left(\frac{|\mathcal{y}_{p_0}(k)|^2}{\frac{1}{P}\sum_{p=1}^{P}|\mathcal{y}_p(k)|^2}\right) \text{dB}. \quad (25)$$

4.2. Experimental Setup

A 3-channel ULA with inter-mic spacing 1 cm centered at [2.4, 2.4, 2.4] m was placed in a $5\times5\times5$ m room with reverberation time $T_{60} = 300$ ms as shown in Fig. 1. Impulse responses were simulated using the the source-image method [13] for $P = 16$ angles on the horizontal plane at radius 1 m from the centre of the array. The look direction was chosen as the endfire steering angle ($p_0 = 180°$). The following parameters were used: sampling frequency $f_s = 8$ kHz, $L = 256$ samples, $L_i = L$ samples. The target response $\mathbf{d}_{p_0}$ was a perfect impulse with delay $\tau = L/2$ samples. The design was solved for the three algorithms under test, then evaluated with (23). The inequality constraint ϵ was set to -20 dB.

4.3. Results and Discussion

The results in Fig. 2 show that MINT introduces the least white noise gain and is therefore least sensitive to sensor noise; this is an intuitive result as the MINT optimization problem is unconstrained and therefore better conditioned than the constrained cases. The FSB result introduces the greatest WNG at all frequencies, with the oracle solution lying in between; this is most likely due to the constant ϵ relaxing the distortionless constraint.

Figs. 3 and 5 show equalized impulse responses stacked across angles and the corresponding DRRs respectively. They reveal that robustness to sensor noise is not correlated with spatial robustness as

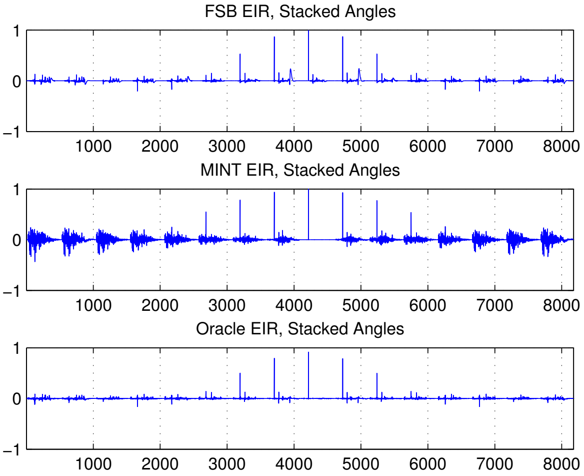

Fig. 3. Equalized impulse responses.

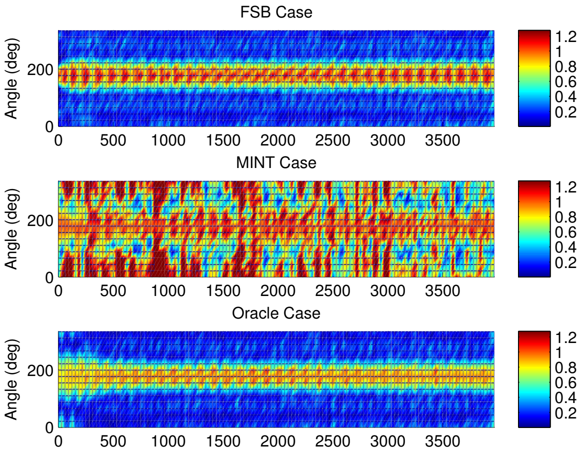

Fig. 4. Reverberant directivity patterns.

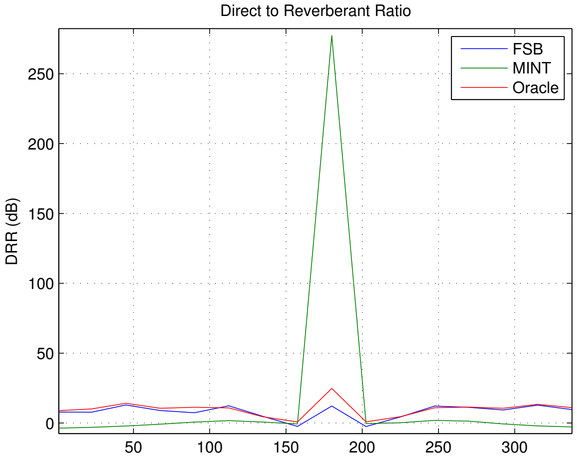

Fig. 5. Direct-to-reverberant ratios.

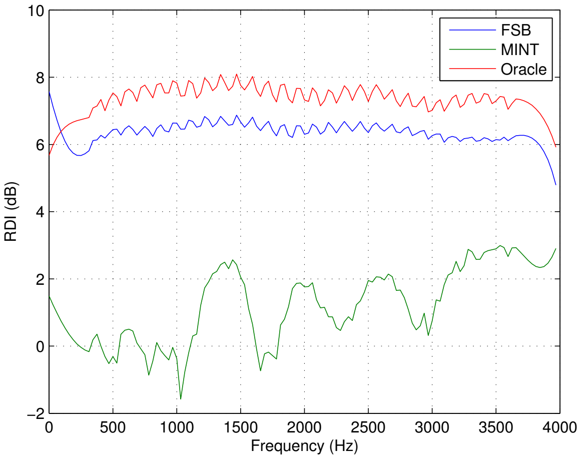

Fig. 6. Reverberant directivity indices.

the MINT solution provides perfect equalization in the look direction and a very poor DRR elsewhere. The oracle solution provides a similar spatial DRR to the FSB but with a 12 dB improvement in the look direction.

Fig. 4 shows the reverberant directivity patterns: the magnitude responses of the equalized systems as a function of frequency and angle under reverberant conditions. The results for the FSB and oracle cases are similar to a classic (anechoic) directivity pattern, showing a main lobe in the look direction but with added distortions due to reverberation. The reverberant directivity pattern is highly chaotic in the MINT case but close inspection reveals unit gain in a thin horizontal stop corresponding to the look direction only. Fig. 6 shows the corresponding reverberant directivity index, highlighting the poor performance of MINT and revealing that the oracle case provides a ~ 1.5 dB improvement over FSB.

These preliminary results suggest that the FSB and MINT paradigms can be combined into a single oracle case that provides the performance in the look direction of MINT and the spatial robustness of a FSB without increasing WNG. The close spacing of the array microphones was chosen to exaggerate WNG; a practical array would most likely have much high inter-microphone spacing with accordingly reduced WNG. These results motivate a deeper investigation into WNG constraints, a wider range of geometries and conditions, the introduction of channel mismatch, variable numbers of control points and comparisons with frequency domain approaches.

5. CONCLUSIONS

Multichannel equalization and optimal beamforming can be formulated as filter and sum operations with differing optimization criteria. Deriving both cases in the time domain, an additional oracle case was proposed that exploits known impulse responses in multiple directions. Under simulated reverberant conditions, it was shown that MINT performs well in the designed look direction but lacks the spatial robustness of the optimal beamformer. Conversely the optimal beamformer lacks channel equalization performance in the look direction. The oracle cases exhibits good properties from both cases without increased sensitivity to sensor noise. This work motivates a deeper investigation under a wider range of conditions.

6. REFERENCES

[1] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag, Berlin, Germany, 2001.

[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] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, pp. 393–398, June 1986.

[4] W. Zhang, E. A. P. Habets, and P. A. Naylor, "A system-identification-error-robust method for equalization of multichannel acoustic systems," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, Mar. 2010.

[5] M. R. P. Thomas, N. D. Gaubitch, and P. A. Naylor, "Application of channel shortening to acoustic channel equalization in the presence of noise and estimation error," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, Oct. 2011.

[6] 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.

[7] J. O. Jungmann, T. Mei, S. Goetze, and A. Mertins, "Room impulse response reshaping by joint optimization of multiple p-norm based criteria," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, Aug. 2011, pp. 1658–1662.

[8] F. Lim, M. R. P. Thomas, and P. A. Naylor, "MINTFormer: A spatially aware channel equalizer," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, Oct. 2013.

[9] E. Mabande, A. Schad, and W. Kellermann, "A time-domain implementation of data-independent robust broadband beamformers with low filter order," Edinburgh, Scotland, May 2011.

[10] C. Lai, S. Nordholm, and Y. Leung, "Design of robust steerable broadband beamformers with spiral arrays and farrow filter structure," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel-Aviv, Israel, Aug. 2010.

[11] S. Yan, Y. Ma, and C. Hou, "Optimal array pattern synthesis for broadband arrays," *J. Acoust. Soc. Am.*, vol. 122, no. 5, pp. 2686–2696, Aug. 2007.

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

[13] 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, Apr. 1979.