DEREVERBERATION SWEET SPOT DILATION WITH COMBINED CHANNEL EQUALIZATION AND BEAMFORMING

M. R. P. Thomas, H. Gamper, I. J. Tashev

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

ABSTRACT

Beamforming and channel equalizers can be formulated as optimal multichannel filter-and-sum operations with different objective criteria. It has been shown in previous studies that the combination of both concepts under a common framework can yield results that combine both the spatial robustness of beamforming and the dereverberation performance of channel equalization. This paper introduces an additional method for leveraging both approaches that exploits channel estimates in a wanted spatial location and derives robustness from knowledge of the array geometry alone. Experiments with an objective assessment of speech quality as a function of source perturbation reveal that the proposed technique can be viewed as a sweet spot dilator when compared with the MINT channel equalizer.

***Index Terms*—** Beamforming, channel equalization, MINT, dereverberation.

1. INTRODUCTION

Speech signals captured by a microphone array at some distance from a wanted talker can be distorted by noise and reverberation, potentially impairing the perceived quality and intelligibility of the recorded signal. Beamforming techniques exploit spatial diversity by filtering and combining the received signals in such a way that suppresses signals incident from directions that are not co-located with the wanted talker, thereby attenuating reverberation components [1]. Beamforming approaches are often classified according to the way in which they estimate and handle the spatial distribution of noise sources, and is an active area of research that has produced many diverse and highly practical results. Many formulations employ a common 'distortionless' constraint to ensure that wanted source is captured without distortion under anechoic conditions, i.e. when the capture vector in the look direction represents pure delays. In addition to other measures to improve robustness, constraints produce solutions that are valid for a wide variety of conditions. Conversely, such beamforming techniques cannot achieve perfect dereverberation as the result is only truly distortionless under anechoic conditions.

Channel equalization techniques adopt a different approach to dereverberation, relying upon *a priori* knowledge of the underlying acoustic channel between a source and an array of microphones. In addition to capturing the propagation time from source to receiver, the ability of the acoustic channels to fully characterize the reverberation of the environment can be exploited. The Multichannel Input/Output Inverse Theorem (MINT) [2] stipulates a set of invertibility conditions that, if satisfied, permit perfect dereverberation of the acoustic environment. Unfortunately, the invertibility conditions are rarely met in practical settings due to mismatch between the true and estimated acoustic channels caused by factors such as modelling error, temperature variations, or movement of the source, receivers, or reflective surfaces [3]. Several variants on the MINT algorithm for channel shortening/reshaping [3, 4, 5] have been proposed, and in some cases additional regularization is applied to improve robustness [6]. MINT-based techniques have also been used to good effect in the practical calibration of microphone arrays [7].

Beamforming and channel equalization can be viewed as optimal filter-and-sum operations with different optimization criteria. Recently, the concept of MINTForming [8] combined frequency domain objective functions to make a parametric tradeoff between the robustness of a superdirective filter-and-sum beamformer (FSB) [9] and the dereverberation performance of MINT. A similar approach formulated a set of time domain constrained optimization problems in which the objective function and constraint could be derived from a filter-and-sum beamformer, MINT, or both [10]. The choice of working in the time domain was to circumvent the FIR approximation required with frequency domain approaches [11, 12, 13] such that perfect dereverberation could be achieved under ideal conditions. In this paper, we introduce an additional 'hybrid' case for situations in which reverberant impulse responses are available only in the direction of the wanted talker, deriving spatial robustness from the array geometry alone. This paper builds upon [10] by considering in details the influence of channel mismatch due to a perturbed source location on speech quality, revealing that the combined FSB/MINT approaches can be viewed as a 'sweet spot dilator' when compared with the pure MINT algorithm.

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, an oracle reference and a hybrid 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. The notation $\check{(\cdot)}$ is used to denote a reverberant impulse response and $\vec{(\cdot)}$ the corresponding anechoic impulse response. Let

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

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

be the impulse responses of length L samples between a source at location $\mathbf{x} = [x, y, z]^T$ and receiver with index $m \in \{1, 2, \ldots, M\}$ in the reverberant and anechoic cases respectively. The source at location $\mathbf{x}_{p_0}$ is considered to be a wanted source; all other locations are modelled as unwanted noise sources. Additionally, let

$$\check{\mathbf{h}}(\mathbf{x}) = [\check{\mathbf{h}}_1^T(\mathbf{x}) \dots \check{\mathbf{h}}_M^T(\mathbf{x})]^T \in \mathbb{R}^{ML \times 1} \quad (3)$$

$$\vec{\mathbf{h}}(\mathbf{x}) = [\vec{\mathbf{h}}_1^T(\mathbf{x}) \dots \vec{\mathbf{h}}_M^T(\mathbf{x})]^T \in \mathbb{R}^{ML \times 1} \quad (4)$$

contain stacked impulse responses between source located at $\mathbf{x}$ and all M microphones. The aim is to synthesize *equalization* 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. Their stacked representation over all microphones is defined in a similar way to (3)

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

Let $\check{\mathbf{H}}_m(\mathbf{x}) \in \mathbb{R}^{(L+L_i-1) \times L_i}$ be a convolution matrix derived from $\check{\mathbf{h}}_m(\mathbf{x})$ so that $\check{\mathbf{H}}_m(\mathbf{x})\mathbf{g}_m$ and $h_m(\mathbf{x}, n) * g_m(n)$ are equivalent, where $*$ denotes linear convolution. A similar formulation is used for $\vec{\mathbf{H}}_m(\mathbf{x})$. Convolution matrices can be stacked for all M channels,

$$\check{\mathbf{H}}(\mathbf{x}) = [\check{\mathbf{H}}_1(\mathbf{x}) \ \cdots \ \check{\mathbf{H}}_M(\mathbf{x})] \in \mathbb{R}^{(L+L_i-1) \times ML_i} \quad (7)$$

$$\vec{\mathbf{H}}(\mathbf{x}) = [\vec{\mathbf{H}}_1(\mathbf{x}) \ \cdots \ \vec{\mathbf{H}}_M(\mathbf{x})] \in \mathbb{R}^{(L+L_i-1) \times ML_i}. \quad (8)$$

In order to formulate a practical design, the source position $\mathbf{x}$ is often quantized into P discrete positions $\mathbf{x}_p$. This yields an additional stacking to form universal convolution matrices:

$$\check{\mathbf{H}} = [\check{\mathbf{H}}(\mathbf{x}_1)^T \dots \check{\mathbf{H}}(\mathbf{x}_p)^T]^T \in \mathbb{R}^{P(L+L_i-1) \times ML_i} \quad (9)$$

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

The response of the equalized system to a source located at $\mathbf{x}_p$ is found by a filter-and-sum operation:

$$\mathbf{y}_p = \sum_{m=1}^{M} \check{\mathbf{H}}_m(\mathbf{x}_p)\mathbf{g}_m = \check{\mathbf{H}}(\mathbf{x}_p)\mathbf{g} \in \mathbb{R}^{(L+L_i-1) \times 1}, \quad (11)$$

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

$$\mathbf{y} = \check{\mathbf{H}}\mathbf{g} \in \mathbb{R}^{P(L+L_i-1) \times 1} \in \mathbb{R}^{P(L+L_i-1) \times 1}, \quad (12)$$

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] stipulates invertibility criteria for unique and exact inverse filtering of room acoustics from multichannel observations:

1. Impulse responses between the wanted source and receivers $\check{\mathbf{h}}_m(\mathbf{x}_{p_0})$ must be known exactly.
2. The convolution matrix $\check{\mathbf{H}}(\mathbf{x}_{p_0})$ is full rank. That is, the individual channels contain no common zeros.
3. The length of the equalization filters L_i satisfies $L_i \geq \left\lceil \frac{L-1}{M-1} \right\rceil$.

Considering only the wanted direction $\mathbf{x}_{p_0}$, MINT equalization filters should produce the equalized output

$$d_{p_0}(l) = \begin{cases} 1 & \text{if } l = \tau; \\ 0 & \text{otherwise}, \end{cases} \quad (13)$$

where τ is an arbitrary delay with vector representation

$$\mathbf{d}_{p_0} = [d_{p_0}(0) \cdots d_{p_0}(L + L_i - 2)]^T \in \mathbb{R}^{(L+L_i-1) \times 1}. \quad (14)$$

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

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \|\check{\mathbf{H}}(\mathbf{x}_{p_0})\mathbf{g} - \mathbf{d}_{p_0}\|_2^2, \quad (15)$$

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

A filter-and-sum beamformer that maximizes ambient noise suppression is one that minimizes the output variance subject to the constraint that sources in the desired location $\mathbf{x}_{p_0}$ are undistorted [1]. In many applications, capture vectors describing the array behavior for a source in location $\mathbf{x}_p$ are pure delays derived from the array geometry. In the time domain the equalizer design can be stated as a convex optimization problem with *distortionless* constraint

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

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

$$\mathbf{d} = [\mathbf{0} \dots \mathbf{d}_{p_0}^T \dots \mathbf{0}]^T \in \mathbb{R}^{P(L+L_i-1) \times 1}. \quad (17)$$

3.3. Oracle Case

Assuming that the reverberant impulse responses $\check{\mathbf{h}}(\mathbf{x}_p)$ are known, a new optimization problem can be defined that fully exploits this information [10]:

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

Notice the constraint in (18) is identical to the MINT objective function in (15). It was proposed in [10] to soften the constraint so as to liberate degrees of freedom to minimize the objective function:

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \|\check{\mathbf{H}}\mathbf{g} - \mathbf{d}\|_2^2 \ \text{subject to} \ \|\check{\mathbf{H}}(\mathbf{x}_{p_0})\mathbf{g} - \mathbf{d}_{p_0}\|_2^2 < \epsilon, \quad (19)$$

where ϵ is an arbitrary constant.

3.4. Hybrid Case

In many cases it is impractical to obtain impulse responses $\check{\mathbf{h}}(\mathbf{x}_p)$ for all design directions as typically only $\check{\mathbf{h}}(\mathbf{x}_{p_0})$ are available. Here we propose a hybrid case that incorporates the anechoic convolution matrix $\vec{\mathbf{H}}$ in the objective function, which can be derived analytically from known array geometry. The constraint incorporates the reverberant convolution matrix in the look direction, $\check{\mathbf{h}}(\mathbf{x}_{p_0})$, in the look direction only,

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

The inequality constant ϵ has been included in a similar fashion to (19). Further details on 3.1–3.3 can be found in [10].

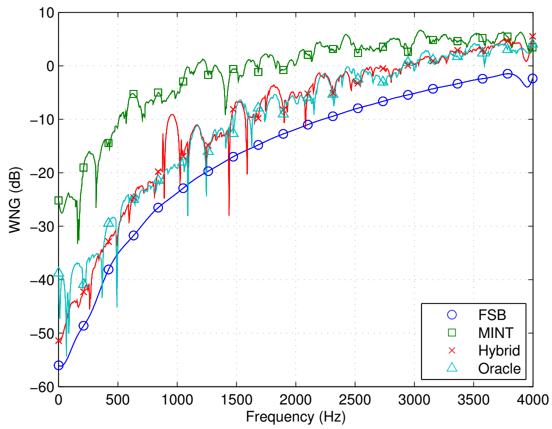

Fig. 1. White noise gains.

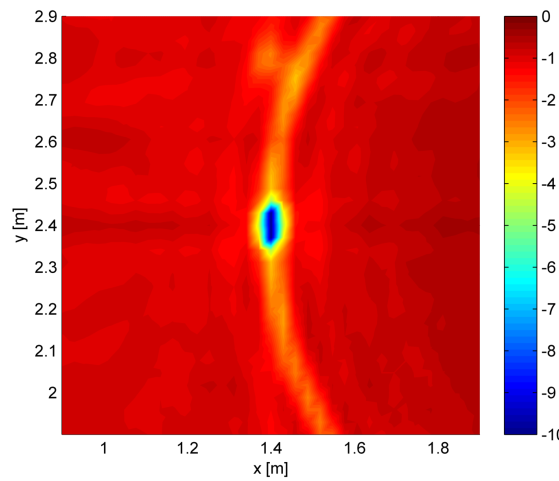

Fig. 2. Normalized projection misalignment (dB) for perturbed source locations.

4. EVALUATION

The experimentation in [10] aimed to evaluate the performance of 3.1–3.3 both in terms of their spatial selectivity and as channel equalizers. They revealed that the oracle case provided both spatial robustness and dereverberation performance in the look direction without increasing white noise gain, and therefore its sensitivity to sensor noise and channel mismatch. While measures such as directivity index [1] and direct-to-reverberant ratio [14] are indicators of performance, the aim of this section is to evaluate equalizer outputs in terms of processed speech quality under noiseless conditions. Specifically, the influence of spatial perturbation of the source in the horizontal plane is considered, such that a mismatch is produced between the channels used in the filter design and the subsequent evaluation.

4.1. Metrics

Normalized projection misalignment [15] is used as a measure of the distance between impulse responses $\check{\mathbf{h}}(\mathbf{x}_{p_0})$ and $\check{\mathbf{h}}(\mathbf{x})$ and has been shown to be aligned with perceptual distance.

$$\text{NPM} = 20\log_{10}\left(\frac{\|\check{\mathbf{h}}(\mathbf{x}_{p_0}) - \eta\check{\mathbf{h}}(\mathbf{x})\|_2}{\|\check{\mathbf{h}}(\mathbf{x}_{p_0})\|_2}\right) \text{ dB} \tag{21}$$

where

$$\eta = \frac{\check{\mathbf{h}}^T(\mathbf{x}_{p_0})\check{\mathbf{h}}(\mathbf{x})}{\check{\mathbf{h}}^T(\mathbf{x})\check{\mathbf{h}}(\mathbf{x})}. \tag{22}$$

White noise gain is defined as

$$\text{WNG} = 10\log_{10}\frac{\left|\mathcal{g}(\omega)^H\check{\mathcal{h}}(\mathbf{x}_{p_0},\omega)\right|^2}{\mathcal{g}(\omega)^H\mathcal{g}(\omega)}, \tag{23}$$

where $\mathcal{g}(\omega) = [\mathcal{g}_1(\omega)\ldots\mathcal{g}_M(\omega)]^T \in \mathbb{C}^{M\times 1}$ and $\check{\mathcal{h}}(\mathbf{x}_{p_0},\omega) = [\check{\mathcal{h}}_1(\mathbf{x}_{p_0},\omega)\ldots\check{\mathcal{h}}_M(\mathbf{x}_{p_0,\omega})]^T \in \mathbb{C}^{M\times 1}$ are vectors of discrete time Fourier transforms of $g_m(n)$ and $\check{h}_m(\mathbf{x}_{p_0},n)$ respectively, and $(\cdot)^H$ is a Hermitian (conjugate) transpose. ITU-T P.862 Perceptual Evaluation of Speech Quality (PESQ) [16] estimates the perceptual speech quality as a predicted mean opinion score (PMOS) in the range -0.5 to 4.5.

4.2. Experimental Setup

A 3-channel uniform linear array with inter-mic spacing 1 cm centered at (2.4, 2.4, 2.4) m was placed in a $5\times 5\times 5$ m room with reverberation time $T_{60} = 600$ ms. Impulse responses were simulated using the the source-image method [17] for $P = 16$ angles on the horizontal plane at radius 1 m from the array array center. The look direction was chosen as the endfire steering angle ($p_0 = 180°$), so that the wanted source lay at (1.4, 2.4, 2.4) m. The following parameters were used: sampling frequency $f_s = 8$ kHz, $L = 2048$ samples, $L_i = L$ samples. The target response $\mathbf{d}_{p_0}$ was a perfect impulse with delay $\tau = L/2$ samples. The inequality constraint ϵ was set to -20 dB. The source position $\mathbf{x}_{p_0}$ was perturbed by up to ± 0.5 m on the horizontal plane.

A speech signal $s(n)$ containing segments of male, female, and children's voices was convolved with the system impulse responses to produce simulated microphone signals $y_m(\mathbf{x},n) = s(n) * \check{h}_m(\mathbf{x},n)$, with corresponding equalizer output

$$\hat{s}(\mathbf{x},n) = \sum_{m=1}^{M} y_m(\mathbf{x},n) * g_m(n). \tag{24}$$

4.3. Results and Discussion

The white noise gains in Fig. 1 reveal that MINT introduces least white noise gain. The hybrid and oracle cases are similar whereas FSB introduces the most. However, it was shown in [10] that while WNG is indicative of robustness to sensor noise, this does not necessarily produce a spatially robust response.

The normalized projection misalignment is shown in Fig. 2 for a source perturbed from the design location of (1.4, 2.4, 2.4) m. Perturbations in the x direction beyond $\pm$ 2 cm rapidly diminish the NPM to over -2 dB. As a rough baseline, informal studies have shown that NPM of better than ~ -8 dB is necessary for good results with MINT using Gaussian channel errors. A ridge of channel errors around -4 dB is a circular arc along which the time of arrival (TOA) of wavefronts at the microphone array are near-constant.

The unprocessed PMOS score at microphone 1 is 2.13. Fig. 3 shows the PMOS results using MINT. At the design location, MINT achieves perfect dereverberation with a corresponding maximum score of 4.5. As the source moves outside a region of $\sim(\pm 5, \pm 15)$

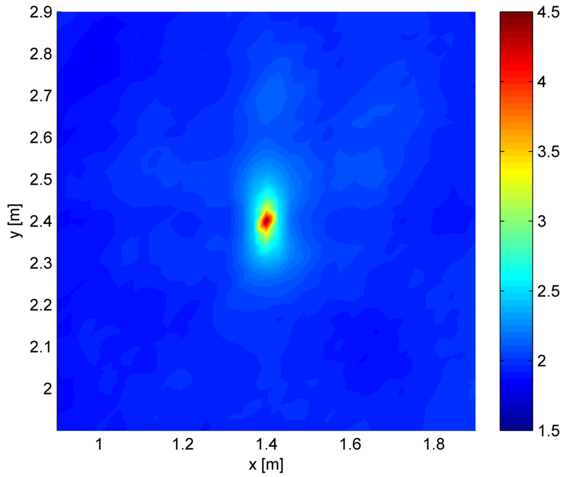

Fig. 3. MINT PMOS Scores.

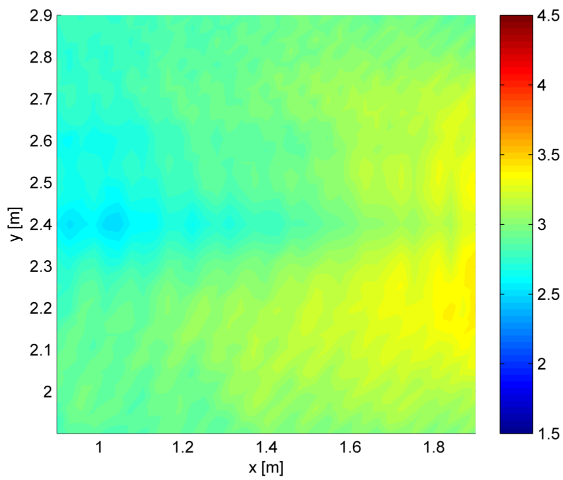

Fig. 4. FSB PMOS Scores.

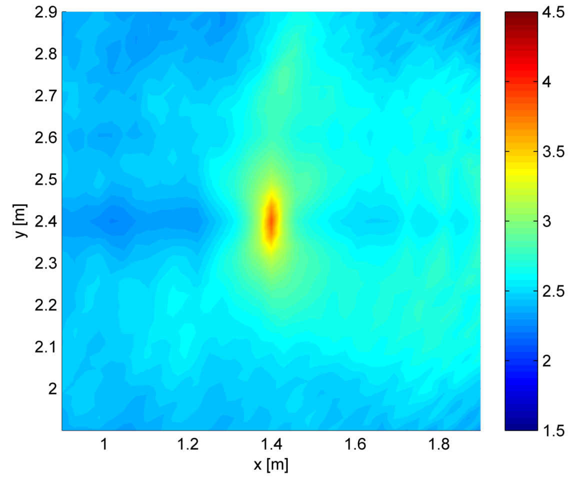

Fig. 5. Oracle PMOS Scores.

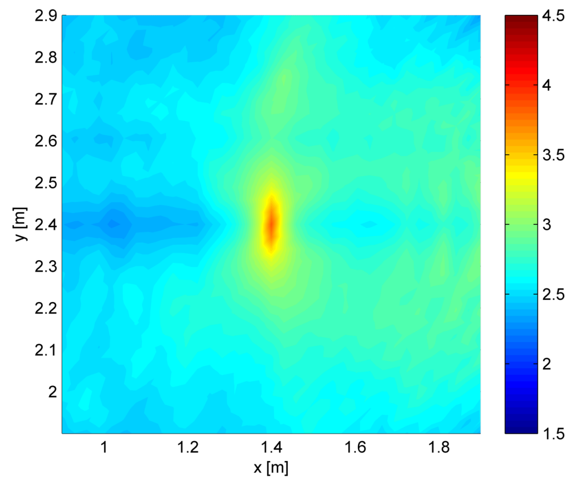

Fig. 6. Hybrid PMOS Scores.

cm, PMOS drops below the unprocessed threshold of 2.13 and into $1.5{-}2$ ('poor'–'bad', 'annoying'–'very annoying'). The area above the PMOS 3 ('fair', 'slightly annoying') is $\sim (\pm 1, \pm 2)$ cm. It is expected that the reason for decreased sensitivity in y is that it lies tangential to the arc of near-constant TOA seen in Fig. 2.

The FSB in Fig. 4 achieves a score of 3-3.5 ('fair', 'slightly annoying') throughout the entire perturbation area. Unlike MINT, there is no maximum at the source design location. Viewed as a main lobe, the perturbation lies in an area where the array gain varies very little, explaining the lack of variation in PMOS. The FSB also relies entirely upon differences in TOAs and not on the reverberant channel; hence it is less sensitive than NPM that considers the entire impulse response. As the source moves closer to the array, PMOS improves due to the improved direct-to-reverberant ratio.

The oracle and hybrid algorithms in Figs. 5 and 6 yield near-identical results, and in neither case does the PMOS quality drop below 2.13. Both remain above PMOS 3 within $\sim (\pm 5, \pm 15)$ cm, peaking at 3.8. Consequently, both schemes can be viewed as a sweet spot dilator to the MINT algorithm. The peak in the oracle case is slightly wider whereas the hybrid case is marginally more spatially robust, although neither exhibits the spatial robustness of the FSB. The similarity of these two sets of results suggests that knowledge of the impulse response towards the wanted direction is more important than for the unwanted directions. It is anticipated that, had ϵ or L_i been increased, the oracle case would start performing better due to increased degrees of freedom with which to make use of the full spatial impulse responses. Experimentation with these parameters, and the behaviour in the presence of noise and white noise gain constraints, are subjects for future study.

5. CONCLUSIONS

The problem of combining channel equalization and beamforming as filter-and-sum operations has been considered. Building upon previous results that combined aspects of both approaches, a new hybrid scheme was proposed that improves the robustness of channel equalization using channel estimates in the wanted direction only. Objective speech quality results reveal that the improved robustness can be viewed as a sweet spot dilator when compared with the MINT channel equalizer, achieving both good dereverberation performance for small perturbations in source location (in a region 5×15 cm), without impairing the result for perturbations up to at least 0.5 m.

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] 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.

[4] 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.

[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] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 9, Sept. 2013.

[7] K. Helwani, H. Buchner, and S. Spors, "Calibration of microphone arrays with aribtrary geometries," in *German Annual Conference on Acoustics (DAGA)*, 2011.

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

[10] M. R. P. Thomas, F. Lim, I. J. Tashev, and P. A. Naylor, "Optimal beamforming as a time domain equalization problem with applications to room acoustics," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Juan Les Pins, France, Sept. 2014.

[11] E. Mabande, A. Schad, and W. Kellermann, "A time-domain implementation of data-independent robust broadband beamformers with low filter order," in *Proc. Hands-Free Speech Communication and Microphone Arrays*, Edinburgh, Scotland, May 2011.

[12] 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.

[13] 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.

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

[15] D. R. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, July 1998.

[16] ITU-T P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," Feb. 2001.

[17] 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.