

ON A TRADEOFF BETWEEN DEREVERBERATION AND NOISE REDUCTION USING THE MVDR BEAMFORMER

Emanuël A. P. Habets^{1,3}, Jacob Benesty², Israel Cohen¹, and Sharon Gannot³

¹ Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
{ehabets, icohen}@ee.technion.ac.il

² INRS-EMT, University of Quebec
Montreal, Quebec, H5A 1K6, Canada
benesty@emt.inrs.ca

³ Bar-Ilan University
Ramat-Gan, 52900, Israel
gannot@eng.biu.ac.il

ABSTRACT

The minimum variance distortionless response (MVDR) beamformer can be used for both speech dereverberation and noise reduction. In this paper we analyse the tradeoff between the amount of speech dereverberation and noise reduction achieved by the MVDR beamformer. We show that the amount of noise reduction that is sacrificed when desiring both speech dereverberation and noise reduction depends on the direct-to-reverberation ratio of the acoustic transfer function between the desired source and a reference microphone. The performance evaluation supports the theoretical analysis and demonstrates the tradeoff between speech dereverberation and noise reduction.

Index Terms— Minimum variance distortionless response (MVDR) filter, noise reduction, speech enhancement, speech dereverberation, microphone arrays, beamforming.

1. INTRODUCTION

Distant or hands-free audio acquisition is required in many applications such as audio-bridging and teleconferencing. Microphone arrays are often used for the acquisition and consist of sets of microphone sensors that are arranged in specific patterns. The received sensor signals usually consist of a desired sound signal, coherent and non-coherent interferences. The received signals are processed in order to extract the desired sound, or in other words to suppress the interferences. In the last four decades many algorithms have been proposed to process the received sensor signals [1, 2].

The minimum variance distortionless response (MVDR) beamformer, also known as Capon beamformer [3], minimizes the output power of the beamformer under a single linear constraint on the response of the array towards the desired signal. The MVDR beamformer can achieve perfect reverberation cancellation when the acoustic transfer functions (ATFs) between the desired source and the microphones are known [2, 4]. In earlier works [4], it was observed that there is a tradeoff between reverberation and noise reduction. However, this tradeoff was never rigorously analysed. In this paper we analyse the tradeoff between noise and reverberation reduction.

2. MVDR BEAMFORMER IN ROOM ACOUSTICS

Consider the conventional signal model in which an N -element sensor array captures a convolved desired signal in some noise field.

This research was supported by the Israel Science Foundation (grant no. 1085/05).

The received signals are expressed as [1]

$$\begin{aligned} y_n(k) &= g_n * s(k) + v_n(k) \\ &= x_n(k) + v_n(k), \quad n = 1, 2, \dots, N, \end{aligned} \quad (1)$$

where g_n is the impulse response from the unknown (desired) source $s(k)$ to the n th microphone, $*$ stands for convolution, and $v_n(k)$ is the noise at microphone n . We assume that the signals $x_n(k)$ and $v_n(k)$ are uncorrelated and zero mean. All signals considered in this work are broadband. Without loss of generality, we consider the first microphone as the reference. *Our main objective is then to study the recovering of anyone of the signals $x_1(k)$ (noise reduction only), a delayed and attenuated version of $s(k)$ (total dereverberation and noise reduction), or a filtered version of $s(k)$ with the MVDR beamformer.*

In the frequency domain, (1) can be rewritten as

$$\begin{aligned} Y_n(j\omega) &= G_n(j\omega)S(j\omega) + V_n(j\omega) \\ &= X_n(j\omega) + V_n(j\omega), \quad n = 1, 2, \dots, N, \end{aligned} \quad (2)$$

where $Y_n(j\omega)$, $G_n(j\omega)$, $S(j\omega)$, $X_n(j\omega) = G_n(j\omega)S(j\omega)$, and $V_n(j\omega)$ are the discrete-time Fourier transforms (DTFTs) of $y_n(k)$, g_n , $s(k)$, $x_n(k)$, and $v_n(k)$, respectively, at angular frequency ω ($-\pi < \omega \leq \pi$) and j is the imaginary unit ($j^2 = -1$).

In vector notation the N microphone signals are given by

$$\mathbf{y}(j\omega) = \mathbf{g}(j\omega)S(j\omega) + \mathbf{v}(j\omega) = \mathbf{x}(j\omega) + \mathbf{v}(j\omega), \quad (3)$$

where

$$\begin{aligned} \mathbf{y}(j\omega) &= [Y_1(j\omega) \quad Y_2(j\omega) \quad \dots \quad Y_N(j\omega)]^T, \\ \mathbf{g}(j\omega) &= [G_1(j\omega) \quad G_2(j\omega) \quad \dots \quad G_N(j\omega)]^T \\ \mathbf{x}(j\omega) &= [X_1(j\omega) \quad X_2(j\omega) \quad \dots \quad X_N(j\omega)]^T, \\ &= S(j\omega)\mathbf{g}(j\omega), \\ \mathbf{v}(j\omega) &= [V_1(j\omega) \quad V_2(j\omega) \quad \dots \quad V_N(j\omega)]^T, \end{aligned}$$

and superscript T denotes transpose of a vector or a matrix.

The beamforming is then performed by applying a complex weight to each sensor and summing across the aperture:

$$Z(j\omega) = \mathbf{h}^H(j\omega)\mathbf{y}(j\omega) = \mathbf{h}^H(j\omega) [\mathbf{g}(j\omega)S(j\omega) + \mathbf{v}(j\omega)], \quad (4)$$

where $Z(j\omega)$ is the beamformer output,

$$\mathbf{h}(j\omega) = [H_1(j\omega) \quad H_2(j\omega) \quad \dots \quad H_N(j\omega)]^T$$

is the beamforming weight vector which is suitable for performing spatial filtering at frequency ω , and superscript H denotes transpose

conjugation of a vector or a matrix.

The PSD of the beamformer output is given by

$$\phi_z(\omega) = \mathbf{h}^H(j\omega)\Phi_x(j\omega)\mathbf{h}(j\omega) + \mathbf{h}^H(j\omega)\Phi_v(j\omega)\mathbf{h}(j\omega), \quad (5)$$

where

$$\Phi_x(j\omega) = E \left[\mathbf{x}(j\omega)\mathbf{x}^H(j\omega) \right] = \phi_s(\omega)\mathbf{g}(j\omega)\mathbf{g}^H(j\omega) \quad (6)$$

is the rank-one PSD matrix of the convolved speech signals with $E(\cdot)$ denoting mathematical expectation, and

$$\Phi_v(j\omega) = E \left[\mathbf{v}(j\omega)\mathbf{v}^H(j\omega) \right] \quad (7)$$

is the PSD matrix of the noise field. In the rest of this paper, we assume that the noise is not fully coherent at the microphones so that $\Phi_v(j\omega)$ is a full-rank matrix.

Now, we define a parameterized desired signal, which we denote by $Q(j\omega)S(j\omega)$, where $Q(j\omega)$ refers to a complex scaling factor that defines the nature of our desired signal. Let $G_1^d(j\omega)$ denote the DTFT of the direct path response from the desired source to the first microphone. By setting $Q(j\omega) = G_1^d(j\omega)$, we are stating that we desire both noise reduction and complete dereverberation. By setting $Q(j\omega) = G_1(j\omega)$, we are stating that we desire only noise reduction.

The MVDR filter can be found by solving the following optimization problem [1]:

$$\begin{aligned} \mathbf{h}_{\text{MVDR}}(j\omega) = \arg \min_{\mathbf{h}(j\omega)} & \mathbf{h}^H(j\omega)\Phi_v(j\omega)\mathbf{h}(j\omega) \\ \text{subject to} & \mathbf{h}^H(j\omega)\mathbf{g}(j\omega) = Q(j\omega). \end{aligned} \quad (8)$$

The solution to this constrained optimization problem is given by

$$\mathbf{h}_{\text{MVDR}}(j\omega) = Q^*(j\omega) \frac{\Phi_v^{-1}(j\omega)\mathbf{g}(j\omega)}{\mathbf{g}^H(j\omega)\Phi_v^{-1}(j\omega)\mathbf{g}(j\omega)}, \quad (9)$$

where superscript $*$ denotes complex conjugation.

3. PERFORMANCE MEASURES

In this section, we present some useful measures that will help us better understand how noise reduction and speech dereverberation work with the MVDR beamformer in a real room acoustic environment.

Since the parameterized desired signal is $Q(j\omega)S(j\omega)$ and the first microphone is the reference signal, we define the local input signal-to-noise ratio (SNR) as

$$\text{iSNR}[Q(j\omega)] = \frac{|Q(j\omega)|^2 \phi_s(\omega)}{\phi_{v_1}(\omega)}, \quad \omega \in (-\pi, \pi], \quad (10)$$

where $\phi_{v_1}(\omega)$ is the PSD of the noise signal $v_1(k)$. The global input SNR is given by

$$\text{iSNR}(Q) = \frac{\int_{-\pi}^{\pi} |Q(j\omega)|^2 \phi_s(\omega) d\omega}{\int_{-\pi}^{\pi} \phi_{v_1}(\omega) d\omega}. \quad (11)$$

After the MVDR beamforming operation with the frequency-domain model given in (5) and using (6), the local output SNR is

$$\begin{aligned} \text{oSNR}[\mathbf{h}_{\text{MVDR}}(j\omega)] &= \frac{|\mathbf{h}_{\text{MVDR}}^H(j\omega)\mathbf{g}(j\omega)|^2 \phi_s(\omega)}{\mathbf{h}_{\text{MVDR}}^H(j\omega)\Phi_v(j\omega)\mathbf{h}_{\text{MVDR}}(j\omega)} \\ &= \phi_s(\omega)\mathbf{g}^H(j\omega)\Phi_v^{-1}(j\omega)\mathbf{g}(j\omega). \end{aligned} \quad (12)$$

It is extremely important to observe that the desired scaling provided by $Q(j\omega)$ has no impact on the resulting local output SNR (but has an impact on the local input SNR). The global output SNR with the MVDR filter is

$$\begin{aligned} \text{oSNR}(\mathbf{h}_{\text{MVDR}}) &= \frac{\int_{-\pi}^{\pi} |\mathbf{h}_{\text{MVDR}}^H(j\omega)\mathbf{g}(j\omega)|^2 \phi_s(\omega) d\omega}{\int_{-\pi}^{\pi} \mathbf{h}_{\text{MVDR}}^H(j\omega)\Phi_v(j\omega)\mathbf{h}_{\text{MVDR}}(j\omega) d\omega} \\ &= \frac{\int_{-\pi}^{\pi} |Q(j\omega)|^2 \phi_s(\omega) d\omega}{\int_{-\pi}^{\pi} \text{oSNR}^{-1}[\mathbf{h}_{\text{MVDR}}(j\omega)] |Q(j\omega)|^2 \phi_s(\omega) d\omega}. \end{aligned} \quad (13)$$

Contrary to the local output SNR, the global output SNR depends strongly on the complex scaling factor $Q(j\omega)$.

Another important measure is the level of noise reduction achieved through beamforming. Therefore, we define the local noise-reduction factor as the ratio of the PSD of the original noise at the reference microphone over the PSD of the residual noise:

$$\begin{aligned} \xi_{\text{nr}}[\mathbf{h}_{\text{MVDR}}(j\omega)] &= \frac{\phi_{v_1}(\omega)}{\mathbf{h}_{\text{MVDR}}^H(j\omega)\Phi_v(j\omega)\mathbf{h}_{\text{MVDR}}(j\omega)} \\ &= \frac{\text{oSNR}[\mathbf{h}_{\text{MVDR}}(j\omega)]}{\text{iSNR}[Q(j\omega)]}, \quad \omega \in (-\pi, \pi]. \end{aligned} \quad (14)$$

The local noise-reduction factor tells us exactly how much the output SNR is improved (or not) compared to the input SNR.

Integrating across the entire frequency range in the numerator and denominator of (14) yields the global noise-reduction factor for the MVDR beamformer:

$$\xi_{\text{nr}}(\mathbf{h}_{\text{MVDR}}) = \frac{\text{oSNR}(\mathbf{h}_{\text{MVDR}})}{\text{iSNR}(Q)}. \quad (15)$$

4. PERFORMANCE ANALYSIS

In this section we analyse the tradeoff between dereverberation and noise reduction.

4.1. On the Comparison of Different MVDR Beamformers

One of the main objectives of this work is to compare MVDR beamformers with different constraints. Obviously, by choosing any constraint $\mathbf{h}^H(j\omega)\mathbf{g}(j\omega) = \gamma \cdot G_1^d(j\omega)$ ($0 < \gamma \leq 1$) we desire both noise reduction and complete dereverberation. The MVDR filters $\mathbf{h}_{\text{MVDR},\gamma}(j\omega)$ are equal to $\gamma \mathbf{h}_{\text{MVDR}}(j\omega)$, i.e., by scaling the desired signal we scale the MVDR filter. When we directly calculate the local noise-reduction factor of the beamformers using (14) we obtain different results since

$$\xi_{\text{nr}}[\gamma \cdot \mathbf{h}_{\text{MVDR}}(j\omega)] \neq \xi_{\text{nr}}[\mathbf{h}_{\text{MVDR},\gamma}(j\omega)]. \quad (16)$$

This can also be explained by the fact that the local output SNRs of all MVDR beamformers $\mathbf{h}_{\text{MVDR},\gamma}(j\omega)$ are equal (since the output SNR is independent of γ) while the local input SNRs are different. A similar problem occurs when we like to compare the noise-reduction factor for MVDR beamformers with completely different constraints because the power of the reverberant signal is much larger than the power of the direct sound signal. This abnormality can be corrected by normalizing the power of the output signal.

To obtain a meaningful noise-reduction factor and to be consistent with earlier works we propose to make the power of the desired signal at the output of the beamformer equal to the power of the signal that would be obtained when using the constraint

$\mathbf{h}^H(j\omega)\mathbf{g}(j\omega) = G_1(j\omega)$. The normalization factor $\eta(Q, G_1)$ is then given by

$$\eta(Q, G_1) = \frac{\int_{-\pi}^{\pi} |Q(j\omega)|^2 \phi_s(\omega) d\omega}{\int_{-\pi}^{\pi} |G_1(j\omega)|^2 \phi_s(\omega) d\omega}. \quad (17)$$

4.2. Local Noise-Reduction Factor

The normalized local noise-reduction factor [as defined in (14)] can be written as

$$\begin{aligned} \tilde{\xi}_{\text{nr}}[\mathbf{h}_{\text{MVDR}}(j\omega)] \\ = \frac{1}{\zeta[Q(j\omega), G_1(j\omega)]} \cdot \frac{\text{oSNR}[\mathbf{h}_{\text{MVDR}}(j\omega)] \cdot \phi_{v_1}(\omega)}{\phi_s(\omega)}, \end{aligned} \quad (18)$$

where $\zeta[Q(j\omega), G_1(j\omega)] = |Q(j\omega)|^2 / \eta(Q, G_1)$. Indeed, for different MVDR beamformers the noise-reduction factor varies due to $\zeta[Q(j\omega), G_1(j\omega)]$, since the local output SNR, $\phi_{v_1}(\omega)$, and $\phi_s(\omega)$ do not depend on $Q(j\omega)$. Since $\zeta[Q(j\omega), G_1(j\omega)] = \zeta[\gamma Q(j\omega), G_1(j\omega)]$ ($0 < \gamma \leq 1$) the normalized local noise-reduction factor is independent of the global scaling factor.

To gain more insight into the local behaviour of $\zeta[Q(j\omega), G_1(j\omega)]$ we analysed several acoustic transfer functions. To simplify the following discussion we assume that the power spectral density $\phi_s(\omega) = 1$ for all ω . Let us decompose the transfer function $G_1(j\omega)$ into two parts. The first part $G_1^d(j\omega)$ is the DTFT of the direct path, while the second part $G_1^r(j\omega)$ is the DTFT of the reverberant part. Now let us define the desired response as

$$Q(j\omega, \alpha) = G_1^d(j\omega) + \alpha G_1^r(j\omega), \quad (19)$$

where the parameter $0 \leq \alpha \leq 1$ controls the direct-to-reverberation ratio (DRR) of the desired response. By setting α to zero we state that we desire complete dereverberation, and by setting α to one we state that we desire no dereverberation. In Fig. 1 we have plotted the probability distribution¹ of $\zeta[Q(j\omega, \alpha), G_1(j\omega)]$ for $\alpha = \{0, 0.2, 1\}$. First, we observe that the probability that $\zeta[Q(j\omega, \alpha), G_1(j\omega)]$ is smaller than the median value of $\zeta[Q(j\omega, 0), G_1(j\omega)]$ decreases when α decreases (i.e., when the DRR increases). Secondly, we observe that the distribution is stretched out towards smaller values (negative values on the logarithmic scale) when α increases. Hence, when the desired speech signal contains less reverberation (i.e., α decreases) it is likely that $\zeta[Q(j\omega, \alpha), G_1(j\omega)]$ will increase and that the local noise-reduction factor will decrease.

4.3. Global Noise-Reduction Factor

Using (15), (13), and (11) we deduce the normalized global noise-reduction factor:

$$\begin{aligned} \tilde{\xi}_{\text{nr}}(\mathbf{h}_{\text{MVDR}}) \\ = \eta(Q, G_1) \frac{\text{oSNR}(\mathbf{h}_{\text{MVDR}})}{\text{iSNR}(Q)} \\ = \frac{\int_{-\pi}^{\pi} \phi_{v_1}(\omega) d\omega}{\int_{-\pi}^{\pi} \text{oSNR}^{-1}[\mathbf{h}_{\text{MVDR}}(j\omega)] \zeta[Q(j\omega), G_1(j\omega)] \phi_s(\omega) d\omega}. \end{aligned} \quad (20)$$

¹The probability distributions were obtained using 50 ATFs (calculated using [5]). The ATFs are obtained under the same room conditions by translating and rotating the source-microphone configuration (i.e, the source-microphone distance is fixed).

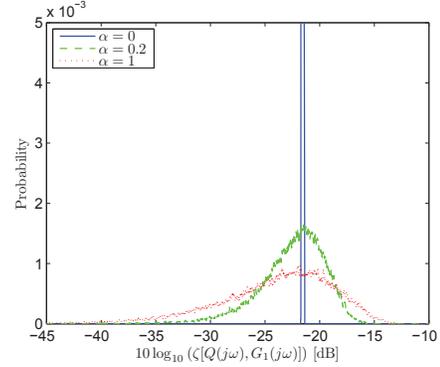


Fig. 1. The probability distribution of $20 \log_{10}(\zeta[Q(j\omega, \alpha), G_1(j\omega)])$ for $\alpha = \{0, 0.2, 1\}$.

This normalized global noise-reduction factor behaves, with respect to $Q(j\omega)$, similarly to its local counterpart. It can easily be verified that the normalized global noise-reduction factor for $\gamma \cdot Q(j\omega)$ is independent of γ . In addition, we observe that only $\zeta[Q(j\omega), G_1(j\omega)]$ depends on $Q(j\omega)$. It is however difficult to predict the exact behaviour of the normalized global noise-reduction factor. From the analyses in the previous subsection we do know that the distribution of $\zeta[Q(j\omega), G_1(j\omega)]$ is stretched out towards smaller values when the DRR decreases. Hence, for each frequency it is likely that $\zeta[Q(j\omega), G_1(j\omega)]$ will decrease when the DRR decreases. Consequently, we expect that the normalized global noise-reduction factor will always increase when the DRR decreases.

5. PERFORMANCE EVALUATION

In this section, we demonstrate the tradeoff between speech dereverberation and noise reduction by computing the normalized noise-reduction factor in various scenarios. A linear microphone array was used with 2 to 8 microphones and an inter-microphone distance of 5 cm. The room size is $5 \times 4 \times 6$ m (length \times width \times height). All room impulse responses are generated using the image-method proposed by Allen and Berkley [5]. The desired source consists of speech like noise (USASI). The noise consists of an AR(1) process (autoregressive process of order one) that was created by filtering a stationary zero-mean Gaussian sequences with a linear time-invariant filter. All signals were analysed using discrete Fourier transforms of length 8192.

In order study the tradeoff more carefully we need to control the amount of reverberation reduction. Here we propose to control the amount of reverberation reduction by changing the DRR of the desired response $Q(j\omega)$. As proposed in Section 4.1, we control the DRR using the parameter α ($0 \leq \alpha \leq 1$) and the complex scaling factor $Q(j\omega, \alpha)$ is calculated using (19).

5.1. Influence of the Number of Microphones

In this section we study the influence of the number of microphones used. The reverberation time was set to $T_{60} = 0.3$ s and the distance between the source and the first microphone was $D = 2$ m. The noise field is non-coherent and the global input SNR [for $Q(j\omega, 0) = G_1^d(j\omega)$] was $\text{iSNR}_{\text{nc}} = 5$ dB. In this experiment 2, 4, or 8 microphones were used. In Fig. 2 the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$. Firstly, we observe that there is a tradeoff between speech dereverberation and noise reduction. The largest amount of noise reduction is achieved for $\alpha = 1$,

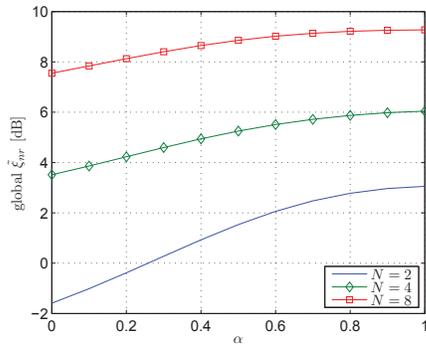


Fig. 2. The normalized global noise-reduction factor obtained using $N = \{2, 4, 8\}$ ($T_{60} = 0.3$, $D = 2$ m, $i\text{SNR}_{nc} = 5$ dB).

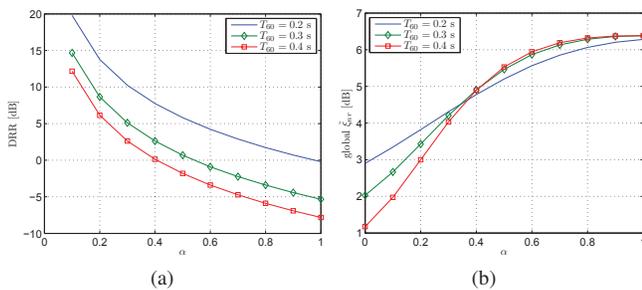


Fig. 3. a) The DRR of $Q(j\omega, \alpha)$ for $T_{60} = \{0.2, 0.3, 0.4\}$ s. b) The normalized global noise-reduction factor obtained using $T_{60} = \{0.2, 0.3, 0.4\}$ s ($N = 4$, $D = 4$ m, $i\text{SNR}_{nc} = 5$ dB).

i.e., when no dereverberation is performed. While a smaller amount of noise reduction is achieved for $\alpha = 0$, i.e., when complete dereverberation is performed. In the case of two microphones ($N = 2$), we amplify the noise when we desire to complete dereverberate the speech signal. Secondly, we observe that the amount of noise reduction increases with approximately 3 dB if we double the number of microphones. Finally, we observe that the tradeoff becomes less evident when more microphones are used. When more microphones are available the degrees of freedom of the MVDR beamformer increases. In such a case the MVDR beamformer is apparently able to perform speech dereverberation without significantly sacrificing the amount of noise reduction.

5.2. Influence of the Reverberation Time

In this section we study the influence of the reverberation time. The distance between the source and the first microphone was set to $D = 4$ m. The noise field is non-coherent and the global input SNR [for $Q(j\omega, 0) = G_1^d(j\omega)$] was $i\text{SNR}_{nc} = 5$ dB. In this experiment four microphones were used, and the reverberation time was set to $T_{60} = \{0.2, 0.3, 0.4\}$ s. The DRR ratio of the desired response $Q(j\omega, \alpha)$ is shown in Fig. 3a. In Fig. 3b the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$. Again, we observe that there is a tradeoff between speech dereverberation and noise reduction. This experiment also shows that almost no noise reduction is sacrificed when we desire to increase the DRR to approximately -5 dB. However, when the DRR is increased further (DRR > -5 dB) the noise-reduction decreases.

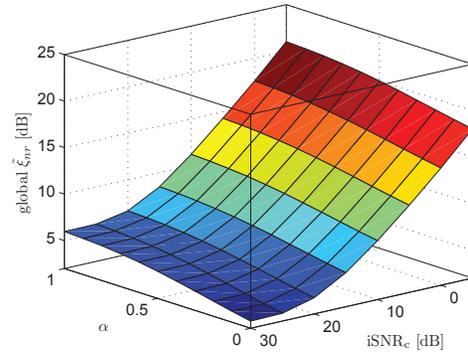


Fig. 4. The normalized global noise-reduction factor obtained using a coherent plus non-coherent noise field $i\text{SNR}_c = \{-5, \dots, 30\}$ dB ($i\text{SNR}_{nc} = 20$ dB, $T_{60} = 0.3$ s, $N = 4$, $D = 2$ m).

5.3. Mixture of a Coherent and Non-Coherent Noise Field

In this section we study the amount of noise reduction in a coherent plus non-coherent noise field with different input SNRs. The input SNR of the non-coherent noise is $i\text{SNR}_{nc} = 20$ dB. The distance between the source and the first microphone was set to $D = 2$ m, $N = 4$, and $T_{60} = 0.3$ s. In Fig. 4 the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$ and for input SNR ($i\text{SNR}_c$) of the coherent noise source that ranges from -5 dB to 30 dB. We observe the tradeoff between speech dereverberation and noise reduction as before. In addition, we see that the noise reduction in a coherent noise field is much larger than the noise reduction in a non-coherent noise field.

6. CONCLUSIONS

We have analysed the tradeoff between speech dereverberation and noise reduction. The performance evaluation supports the theoretical analysis and demonstrates the tradeoff between speech dereverberation and noise reduction. The amount of noise reduction that is sacrificed when complete dereverberation is required depends on the direct-to-reverberation ratio of the acoustic impulse response between the source and the reference microphone. When desiring both speech dereverberation and noise reduction the results also demonstrate that the amount of noise reduction that is sacrificed decreases when the number of microphones increases.

7. REFERENCES

- [1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.
- [2] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2007, ch. 48.
- [3] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
- [4] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1053–1065, Mar. 2007.
- [5] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.