ANALYSIS OF THE DEREVERBERATION PERFORMANCE OF MICROPHONE ARRAYS

Nikolay D. Gaubitch, Patrick A. Naylor

ndg@imperial.ac.uk
Department of Electrical and Electronic Engineering
Imperial College London
SW7 2AZ, UK

ABSTRACT

Microphone arrays and beamforming have been used extensively for speech and audio acquisition in noisy and reverberant environments. In this paper, our objective is to investigate analytically the dereverberation performance of a beamformer. As an evaluation metric we apply the commonly used direct-to-reverberant ratio (DRR). Consequently, we utilize tools from statistical room acoustic theory to derive an expression for the expected DRR in the case of a delay-and-sum beamformer (DSB). Simulation results are provided to demonstrate and validate the theoretical expression.

1. INTRODUCTION

Microphone arrays are proposed for use in many signal processing applications such as speech enhancement and blind source separation [1] with the inherent spatial information as a main advantage over a single microphone. A fundamental multi-microphone technique is steered beamforming, where the mth microphone is steered towards the source using steering filters, $G_m(e^{j\omega})$, and response shaping weights, w_m, as shown in Fig. 1. Here, we focus particularly on the performance of a beamformer for sound acquisition in reverberant environments. This is one area in which beamformers have been demonstrated to be useful with several proposed algorithms [1, 2, 3].

The choice of evaluation metric for enhanced reverberant speech is still an unresolved issue. There has been plethora of research on the topic of intelligibility of music and speech in reverberant environments in the fields of psychoacoustics and architectural acoustics, where the existing evaluation methods are mainly based on the room impulse response [4]. Most of these approaches consist of a ratio between some portion of the direct and the early reflections and the whole impulse response including or excluding the direct component. In general, reflections in the region of $50 - 100$ ms are considered to have a less detrimental effect on the intelligibility. However, for speech in particular, any reverberation impairs the intelligibility of speech to some extent [4].

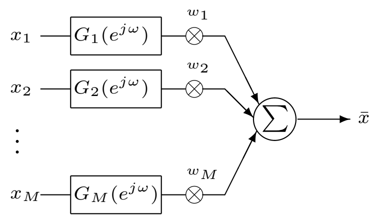

Figure 1: *A general beamformer structure.*

In this paper, we derive an expression for the expected beamformer performance for dereverberation. For the purpose of analysis, we use the delay-and-sum beamformer (DSB) as an example. The DSB is the simplest beamformer where the input signals are aligned so that the direct paths will add constructively. It is of interest also since it has been used as a benchmark for several newly proposed dereverberation algorithms, e.g. [5, 6, 7]. We employ the direct-to-reverberant (DRR) ratio as a performance metric. The DRR measure is defined as the ratio of the energy of the direct signal component only to the energy of the reflected signal components, excluding the direct component.

We use tools from statistical room acoustics (SRA) in order to find the expected improvement in DRR at the beamformer output compared to the best microphone, which is normally the microphone closest to the source, i.e.

$$E\{\hat{\gamma}\} = 10\log_{10}\left(\frac{E\{\gamma_{DSB}\}}{E\{\gamma'\}}\right), \tag{1}$$

where $E\{\cdot\}$ is the spatial expectation, γ' is the DRR of the microphone closest to the source and γ_{DSB} is the DRR at the output of the DSB.

SRA theory provides a statistical model of the room transfer function allowing for a mathematically tractable analysis of highly complex acoustic environments [4]. This diffuse sound field model is considered to closely represent the acoustic properties of rooms provided that:

i) the room dimensions are large compared to the wavelengths of interest, ii) the average spacing between resonant frequencies is smaller than one third of their bandwidth (which is normally considered to be above the Schroeder frequency, $f_{Sch} = 2000\sqrt{T_{60}/V}$ Hz, where T_{60} is the reverberation time and V is the room volume) and iii) speakers and microphones are at least a half wavelength away from the surrounding walls. The statistical room model has been used previously to analyze several signal processing techniques. Radlović et al. [8] and Talantzis and Ward [9] analyzed the robustness of equalizers in reverberant environments, Ward [10] investigated the performance of cross-talk cancellation in reverberant rooms and most recently, Bharitkar et al. used SRA to analyze the performance of spatially averaged room transfer functions for equalization [11] to name a few.

The remainder of this paper is organized as follows. In Section 2, the theoretical expression for the the expected DRR improvement of the DSB is derived, which then enables the analysis of the performance of the beamformer for dereverberation. Section 3 presents experiments and simulation results, provided to validate the theory and to demonstrate the performance of a particular microphone setup. Finally, in Section 4 conclusions are drawn from this work.

2. PERFORMANCE ANALYSIS

The main theoretical result in this paper is summarized in the following Equation, which gives the expected improvement in direct-to-reverberant ratio that can be achieved with a DSB:

$$E\{\hat{\gamma}\} = 10\log_{10}(\Xi), \qquad (2)$$

with

$$\Xi = \frac{D'^2 \sum_{m=1}^{M}\sum_{n=1}^{M} 1/D_m D_n}{\sum_{m=1}^{M}\sum_{n=1}^{M} \frac{\sin k\|\boldsymbol{\ell}_m-\boldsymbol{\ell}_n\|}{k\|\boldsymbol{\ell}_m-\boldsymbol{\ell}_n\|}\cos(k[D_m-D_n])},$$

where D_m is the distance between the source and the mth microphone, $D' = \min(D_m)$ is the distance from the source to the closest microphone and $\boldsymbol{\ell}_m$ is the mth microphone three dimensional coordinate vector. The wave number is $k = 2\pi f/c$ with f denoting the frequency and c being the speed of sound in air which we take as $c = 344$ m/s in room temperature.

The following observations can be made from the expression in (2): i) the expected improvement that can be achieved with the DSB depends only on the distance between the source and the array and the separation of the microphones, ii) consequently, the improvement is independent of the reverberation time and iii) in the special case, when the microphones are separated by exactly a half wavelength at each frequency and the distance between the source and the microphones is large, the denominator tends to zero and the improvement is infinite, i.e. perfect dereverberation is achieved. In the remainder of this section the main steps of the derivation of (2) will be shown.

Consider an array of $M > 1$ microphones with steering filters, $G_m(e^{j\omega})$, and response shaping weights, w_m, as shown in Fig. 1. In the case of the DSB, the weights are uniform, i.e. $w_m = 1/M$, and the steering filters are $G(e^{j\omega}) = e^{-j2\pi f\tau_m}$, where τ_m is a delay to compensate for propagation delays of the mth direct path. Consequently, the transfer function at the output of the DSB is

$$\bar{H}(e^{j\omega}) = \frac{1}{M}\sum_{m=1}^{M} H_m(e^{j\omega})e^{-j2\pi f\tau_m}, \qquad (3)$$

where $H_m(e^{j\omega})$ is the room transfer function from the source to the mth microphone.

In SRA it is assumed that the room transfer function of the mth microphone can be expressed in terms of a direct and a reverberant component, i.e.

$$H_m(e^{j\omega}) = H_{d,m}(e^{j\omega}) + H_{r,m}(e^{j\omega}). \qquad (4)$$

It is further assumed that the direct and reverberant components are uncorrelated and thus, the spatial expectation of the cross terms arising in the squared magnitude of the transfer function are zero. Consequently, the expected energy density spectrum can be written

$$E\{|H_m(e^{j\omega})|^2\} = |H_{d,m}(e^{j\omega})|^2 + E\{|H_r(e^{j\omega})|^2\}, \qquad (5)$$

The direct component, $H_{d,m}(e^{j\omega})$, is given by [8]

$$H_{d,m}(e^{j\omega}) = \frac{e^{jkD_m}}{4\pi D_m} \qquad (6)$$

and from the statistical room response model, the reverberant component, $E\{|H_r(e^{j\omega})|^2\}$, can be written [8]

$$E\{|H_r(e^{j\omega})|^2\} = \left(\frac{1-\alpha}{\pi A\alpha}\right), \qquad (7)$$

where α is the average wall absorption coefficient and A is total wall surface area. Also, it can be shown that the spatial correlation of the reverberant components between the mth and the nth microphones is [10]

$$E\{H_{r,m}(e^{j\omega})H^*_{r,n}(e^{j\omega})\} = \left(\frac{1-\alpha}{\pi A\alpha}\right)\frac{\sin k\|\boldsymbol{\ell}_m-\boldsymbol{\ell}_n\|}{k\|\boldsymbol{\ell}_m-\boldsymbol{\ell}_n\|}, \qquad (8)$$

where $(\cdot)^*$ denotes the complex conjugate.

Using (6) and (7) the expected DRR of the mth microphone can be written [8]

$$\begin{aligned} E\{\gamma_m\} &= \frac{|H_{d,m}(e^{j\omega})|^2}{E\{|H_r(e^{j\omega})|^2\}} \\ &= \frac{\alpha A}{16\pi D_m^2(1-\alpha)}. \end{aligned} \qquad (9)$$

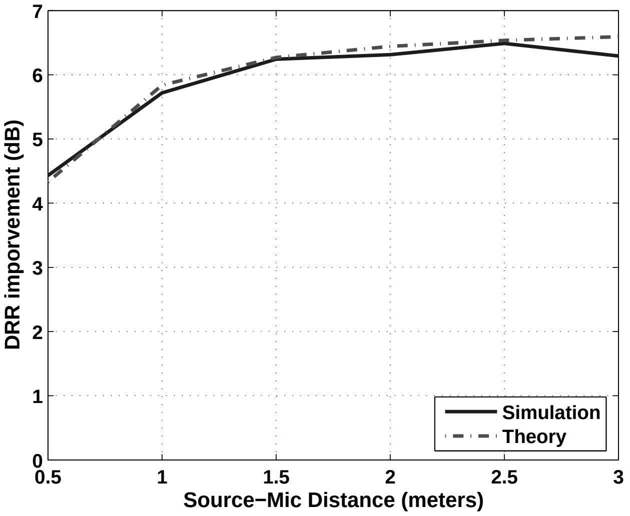

Figure 2: *DRR improvement vs. source-microphone distance for an array of $M = 5$ microphones.*

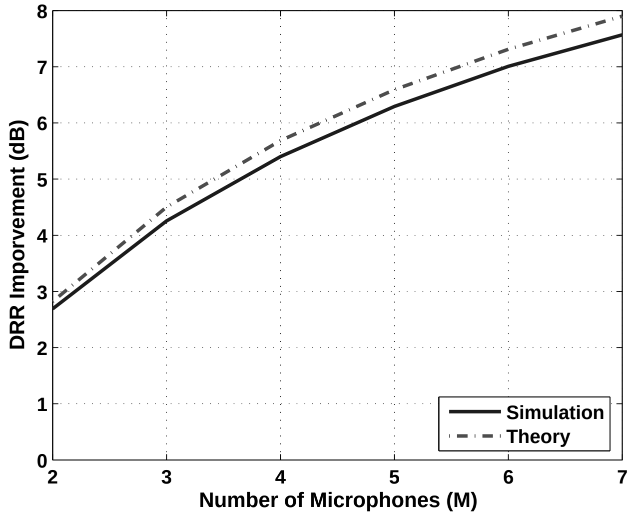

Figure 3: *DRR improvement vs. number of microphones at a fixed distance of $D' = 2$ m.*

Similarly, the expected power density spectrum of the beamformer can be expressed in terms of a direct and reverberant component as in (5), such that $E\{|\bar{H}(e^{j\omega})|^2\} = |\bar{H}_d(e^{j\omega})|^2 + E\{|\bar{H}_r(e^{j\omega})|^2\}$, where the direct component becomes

$$\begin{aligned} |\bar{H}_d(e^{j\omega})|^2 &= \left| \frac{1}{M} \sum_{m=1}^{M} H_{d,m}(e^{j\omega}) e^{-j2\pi f \tau_m} \right|^2 \\ &= \frac{1}{(4\pi M)^2} \sum_{m=1}^{M} \sum_{n=1}^{M} \frac{1}{D_m D_n}, \end{aligned} \tag{10}$$

and the reverberant component can be shown to be

$$E\{|\bar{H}_r(e^{j\omega})|^2\} = \left| \frac{1}{M} \sum_{m=1}^{M} H_{r,m}(e^{j\omega}) e^{-j2\pi f \tau_m} \right|^2 \tag{11}$$

$$= \left(\frac{1-\alpha}{M^2 \pi A \alpha} \right) \sum_{m=1}^{M} \sum_{n=1}^{M} \frac{\sin k\|\boldsymbol{\ell}_m - \boldsymbol{\ell}_n\|}{k\|\boldsymbol{\ell}_m - \boldsymbol{\ell}_n\|} \cos(k[D_m - D_n]),$$

where we have set $\tau_m = D_m/c$ as in [9].

We obtain the expected direct-to-reverberant ratio of the DSB by

$$E\{\gamma_{DSB}\} = \frac{|\bar{H}_d(e^{j\omega})|^2}{E\{|\bar{H}_r(e^{j\omega})|^2\}}, \tag{12}$$

where $|\bar{H}_d(e^{j\omega})|^2$ and $E\{|\bar{H}_r(e^{j\omega})|^2\}$ are defined in (10) and (11) respectively.

Finally, substituting (12) and (9) into (1), we obtain the result stated in (2).

3. SIMULATIONS

In this section, experiments and simulation results are presented to validate the theoretical expression in (2) and also to gain some insight in the expected performance of the DSB for dereverberation.

For the simulations, the source image method [12] was used to generate finite room impulse responses, $h_m(n)$. The room transfer function, $H_m(e^{j\omega})$, was then found by taking the Fourier transform of $h_m(n)$. A room with the dimensions $4 \times 5 \times 6.4$ m was modeled choosing the dimensions as in [8] and [9] so that the best approximation of a diffuse sound field is achieved. A linear array of M microphones with the spacing between adjacent microphones $\|\boldsymbol{\ell}_m - \boldsymbol{\ell}_{m+1}\| = 0.2$ m. The reverberation time was set to $T_{60} = 0.5$ s giving $\alpha = 0.2656$. Frequencies between $300 - 3400$ Hz were considered and sources and microphones were kept at least a half wavelength away from the walls, to satisfy the conditions set for the statistical room model.

To simulate the spatial expectation, $E\{\cdot\}$, we used the approach adopted in [8]. The microphone array and source configuration was kept fixed while rotated and translated to $N = 100$ random positions in the room. In this way, the distance between source and microphones and the microphone separation is kept constant. At each position the direct component was subtracted from the impulse response and the improvement in direct-to-reverberant ratio was calculated. The final result was obtained by averaging the results from the N different positions.

In our first experiment we used an array with $M = 5$

microphones and gradually increased the distance between the array and the source from 0.5 to 3 m. The distance from the source to the array is defined here as the distance to the closest microphone. The result is shown in Fig. 2, where the improvement in DRR calculated with the expression in (2) is plotted with a dashed line while the experimental result is shown with a solid line. It can be observed here that the improvement increases slightly with the distance but then flattens out. This can be related to the theoretical expression by noting that the improvement is mainly governed by the microphone separation when the distance to the array is large.

For the second experiment, the distance between the source and the array was kept fixed at 2 m while the number of microphones was increased. The result of this is shown in Fig. 3, where the improvement in DRR calculated with the expression in (2) is plotted with a dashed line and the experimental result is shown with a solid line. Here it can be seen that the improvement is proportional to the number of microphones.

In both experiments it is clearly demonstrated that the results from the simulations closely match those calculated from theory.

4. CONCLUSIONS

In this paper, we have used statistical room acoustic theory to derive an expression for the expected direct-to-reverberant ratio improvement that can be achieved with a delay-and-sum beamformer in reverberant rooms. From this expression it could be seen that the relative improvement depends only on the microphone spacing and on the distance of the source from the array, where the effect of the latter decreases as the distance is increased. Thus, for a given geometric setup the DRR improvement is independent of the reverberation time. Simulation results were presented to confirm the validity of the derived expression. The theoretical result in this paper together with the DRR measure could thus be applied as a benchmark for new multi-microphone dereverberation algorithms and also for further analysis of different microphone array configurations for the DSB in reverberant environments.

5. REFERENCES

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

[2] J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Amer.*, vol. 62, no. 4, pp. 912–915, Oct. 1977.

[3] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Amer.*, vol. 78, no. 5, pp. 1508–1518, Nov. 1985.

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

[5] N. D. Gaubitch, P. A. Naylor, and D. B. Ward, "Mulit-microphone speech dereverberation using spatio-temporal averaging," in *Proc. 12th European Signal Processing Conf. (EUSIPCO-04)*, Vienna, Austria, Sept. 2004, pp. 809–812.

[6] B. W. Gillespie, H. S. Malvar, and D. A. F. Florêncio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2001, vol. 6, pp. 3701–3704.

[7] S. M. Griebel and M. S. Brandstein, "Wavelet transform extrema clustering for multi-channel speech dereverberation," in *IEEE Workshop on Acoustic Echo and Noise Control*, Pocono Manor, Pennsylvania, Sept. 1999.

[8] B. D. Radlović, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," vol. 8, no. 3, pp. 311–319, May 2000.

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

[10] D. B. Ward, "On the performance of acoustic crosstalk cancellation in a reverberant environment," *J. Acoust. Soc. Amer.*, vol. 110, no. 2, pp. 1195–1198, Aug. 2001.

[11] 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.

[12] 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.