TOWARDS MULTI-MICROPHONE SPEECH DEREVERBERATION USING SPECTRAL ENHANCEMENT AND STATISTICAL REVERBERATION MODELS

Emanuël A. P. Habets[†]

Department of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel

School of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel

ABSTRACT

In speech communication systems the received microphone signals are degraded by room reverberation and ambient noise. Reverberant speech can be separated into two components, viz. an early speech component and a late reverberant speech component. In this paper a multichannel dereverberation algorithm is proposed to suppress late reverberation. Specifically, we employ a minimum variance distortionless beamformer and a single-channel MMSE estimator, which operates on the beamformer's output signal. The so-called late reverberant spectral variance (LRSV) required by the MMSE estimator can be estimated using i) the beamformer's output signal or ii) the received microphone signals. In this contribution we investigate both approaches and show how *a priori* knowledge of the reverberant sound field can be exploited to improve the LRSV estimation. Advantages and disadvantages of the LRSV estimators are discussed, and experimental results using simulated reverberant speech are presented.

Index Terms— Speech dereverberation, minimum variance distortionless response (MVDR) beamformer, minimum mean square error (MMSE) estimator.

1. INTRODUCTION

Distant or hands-free audio acquisition is required in many applications such as audio-bridging and teleconferencing. Microphone arrays can be 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 and interferences such as room reverberation and ambient noise. The degradation of the desired sound caused by these interferences decrease the fidelity and intelligibility of the desired speaker. The received signals are processed in order to extract the desired sound, or in other words to suppress the interferences.

Reverberation reduction methods are generally divided into two categories. Methods of the first category are known as reverberation cancellation methods. In general, a linear filter operation is applied to the observed microphone signals to obtain an estimate of the anechoic signal. The filters are either estimated directly from the observed signals or indirectly using an estimate of the acoustic impulse responses (AIRs) of the acoustic channels between the source and the microphones. Methods in the second category are known as reverberation suppression methods. These methods commonly apply a non-linear operation to the observed microphone signals to suppress reverberation and require little or no *a priori* knowledge about the AIRs. In both categories single and multiple microphone signals are exploited. A major advantage of the suppression methods is the robustness with respect to changes in the room and position changes of the source.

Here we focus on a multi-microphone reverberation suppression method. Specifically, we employ a minimum variance distortionless response (MVDR) beamformer and a single-channel MMSE estimator, which operates on the beamformer's output signal. The single-channel MMSE estimator requires an estimate of the short-term power spectral density (or in the context of statistical spectral enhancement methods, spectral variance) of the late reverberant speech component. In the last decade several late reverberant spectral variance (LRSV) estimators have been developed. In [1] Lebart et al. developed a single-microphone LRSV estimator. In [2] it is shown that the LRSV estimate can be improved by using multiple microphone signals. It should be noted that both estimators implicitly assume that the source-receiver distance is larger than the so-called critical distance. A dual-microphone LRSV estimator which is able to compensate for the energy of the direct sound in case the source-array distance is smaller than the critical distance was first developed in [3]. In [4] a generalized statistical reverberation model was proposed and utilized to develop an estimator which is advantageous in case the source-receiver distance is smaller than the critical distance. A more comprehensive overview of these statical reverberation models and LRSV estimators can also be found in [5].

Here we need to estimate the spectral variance of the residual late reverberant signal at the output of the MVDR beamformer. When multiple microphone are available the LRSV can be estimated using i) the output signal of the MVDR beamformer or ii) the received microphone signals. In this contribution we investigate both approaches and show how *a priori* knowledge of the reverberant sound field can be exploited to improve the LRSV estimation when the received microphone signals are used. Specifically, we show how the spatial coherence of the reverberant sound field can be used to improve the estimate of the LRSV. Advantages and disadvantages of the single- and multi-microphone LRSV estimators are discussed. Experimental results obtained using simulated reverberant speech are presented and illustrate the advantage of the multi-microphone LRSV estimator.

The paper is organized as follows: In Section 2 the problem is formulated. In Section 3 we describe the proposed multichannel dereverberation algorithm. In Section 4 we briefly review the single-microphone LRSV estimator and propose a novel multi-microphone LRSV estimator. Experimental results that demonstrate the performance of both LRSV estimators are presented in Section 5. Finally, conclusions are provided in Section 6.

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

2. PROBLEM FORMULATION

The reverberant signal at the mth microphone results from the convolution of the anechoic speech signal $s(n)$ and a causal AIR $h_m(n)$. Here we assume that the AIR is time-invariant and that its length is infinite. The reverberant speech signal at discrete-time n can be written as

$$z_m(n) = \sum_{n'=0}^{\infty} h_m(n')\, s(n-n'). \tag{1}$$

The mth microphone signal is given by

$$x_m(n) = z_m(n) + v_m(n), \tag{2}$$

where $v_m(n)$ denote the additive ambient noise received by the mth microphone.

In the short-time Fourier transform (STFT) domain we can express the microphone signal as

$$X_m(\ell,k) = \sum_{k'=0}^{K-1} \sum_{\ell'=-\infty}^{\infty} H_m(\ell',k,k')\, S(\ell-\ell',k') + V_m(\ell,k) \tag{3}$$

where ℓ denotes the time index, k and k' denote the band and cross-band frequency indices, K denotes the number of frequency bins, and $V_m(\ell,k)$ denotes the spectral noise component of the mth microphone signal. The STFT response $H(\ell',k,k')$ may be interpreted as a response to an impulse $\delta(\ell',k-k')$ in the time-frequency domain.

To simplify the following discussion, and without loss of generality, it is assumed that the direct sound arrives at time instance n at the first microphone. Since our objective is to suppress late reverberation we split the AIR into two components such that

$$H_m(\ell,k,k') = \begin{cases} 0 & \text{for } \ell < 0; \\ H_{\mathrm{e},m}(\ell,k,k') & \text{for } 0 \le \ell < N_\mathrm{e}; \\ H_{\ell,m}(\ell,k,k') & \text{for } N_\mathrm{e} \le \ell \le \infty, \end{cases} \tag{4}$$

where $H_{\mathrm{e},m}(\ell,k,k')$ models the direct path and a few early reflections and $H_{\ell,m}(\ell,k,k')$ models all later reflections, and N_e ($N_\mathrm{e} \ge 1$) controls the time instance (measured with respect to the arrival time of the direct sound) that indicates the beginning of late reverberation. This parameter can be determined according to design specification of the system or controlled by the listener depending on his or her subjective preference.

Using (4) we can write the microphone signal $X(\ell,k)$ as

$$X_m(\ell,k) = \sum_{k'=0}^{K-1} \sum_{\ell'=-\infty}^{N_\mathrm{e}-1} H_m(\ell',k,k')\, S(\ell-\ell',k') + \sum_{k'=0}^{K-1} \sum_{\ell'=N_\mathrm{e}}^{\infty} H_m(\ell',k,k')\, S(\ell-\ell',k') + V_m(\ell,k). \tag{5}$$

We can write (5) as

$$X_m(\ell,k) = Z_{\mathrm{e},m}(\ell,k) + Z_{\ell,m}(\ell,k) + V_m(\ell,k), \tag{6}$$

where $Z_{\mathrm{e},m}(\ell,k)$ and $Z_{\ell,m}(\ell,k)$ denote the early and late spectral speech components at the mth microphone, respectively.

Our objective is to obtain an estimate of the early speech component without using detailed knowledge of the AIRs. Instead of estimating $Z_{\mathrm{e},m}(\ell,k)$ with $m \in \{1,\dots,M\}$ we propose to estimate a spatially filtered version of all early speech components.

3. MULTICHANNEL REVERBERATION SUPPRESSOR

The proposed multichannel reverberation suppression algorithm consists of two stages. First, a MVDR beamformer is applied to the microphone signals. Second, a single-channel MMSE estimator is applied to the output of the MVDR beamformer. These stages are described in the succeeding subsections.

3.1. MVDR beamformer

Let us define $\mathbf{X}(\ell,k) = [X_1(\ell,k), X_2(\ell,k), \dots, X_M(\ell,k)]^T$ and $\mathbf{V}(\ell,k) = [V_1(\ell,k), V_2(\ell,k), \dots, V_M(\ell,k)]^T$. The MVDR filter is found by solving the following minimization problem

$$\mathbf{W}_{\mathrm{MVDR}}(\ell,k) = \underset{\mathbf{W}(\ell,k)}{\operatorname{argmin}} \left\{ (\mathbf{W}(\ell,k))^H\, \mathbf{\Lambda}_{\mathbf{XX}}(\ell,k) \mathbf{W}(\ell,k) \right\} \quad \text{subject to } (\mathbf{W}(\ell,k))^H\, \mathbf{C}(k) = 1, \tag{7}$$

where $(\cdot)^H$ denotes the Hermitian transpose, $\mathbf{W}(\ell,k) = [W_1(\ell,k), W_2(\ell,k), \dots, W_M(\ell,k)]^T$, $\mathbf{\Lambda}_{\mathbf{XX}}(\ell,k) = E\{\mathbf{X}(\ell,k)\mathbf{X}^H(\ell,k)\}$, and $\mathbf{C}(k)$ denotes a pre-defined constraint column vector of length M.

A major question remains how to define the constraint $\mathbf{C}(k)$ and thereby the signal which is undistorted by the MVDR beamformer. One solution would be to estimate the reverberant speech component $Z_m(\ell,k)$ for $m \in \{1,\dots,M\}$ (see for example [6]). In this case the beamformer only reduces noise (and therefore no reverberation). Here we chose to align the direct sound signals of the desired source. Due to the spatial directivity of the beamformer the spectral coloration induced by early reflections is slightly reduced.

Let us assume that the desired source is in the far-field, such that the propagation of the direct sound can be modelled by $\mathbf{H}_\mathrm{d}(k) = [1, \mathrm{e}^{-j\omega_k \tau_2}, \dots, \mathrm{e}^{-j\omega_k \tau_M}]^T$, where $\omega_k = 2\pi f_\mathrm{s} k/K$, f_s denotes the sampling frequency, and τ_m denotes the time difference of arrival (TDOA) of the desired speech signal at the mth microphone and the first microphone. In this case the constraint is given by

$$\mathbf{C}(k) = \mathbf{H}_\mathrm{d}(k). \tag{8}$$

Estimation of the TDOAs is beyond the scope of this paper in which we assume that the TDOAs are known. The solution of the MVDR beamformer is now given by

$$\mathbf{W}_{\mathrm{MVDR}}(\ell,k) = \frac{\mathbf{\Lambda}_{\mathbf{VV}}^{-1}(\ell,k)\mathbf{H}_\mathrm{d}(k)}{\mathbf{H}_\mathrm{d}^H(k)\mathbf{\Lambda}_{\mathbf{VV}}^{-1}(\ell,k)\mathbf{H}_\mathrm{d}(k)}, \tag{9}$$

where $\mathbf{\Lambda}_{\mathbf{VV}}(\ell,k) = E\{\mathbf{V}(\ell,k)\mathbf{V}^H(\ell,k)\}$. Here we assume that the noise covariance matrix $\mathbf{\Lambda}_{\mathbf{VV}}(\ell,k)$ is estimated during noise-only periods. The output of the MVDR beamformer is given by

$$\begin{aligned} Q(\ell,k) &= (\mathbf{W}_{\mathrm{MVDR}}(\ell,k))^H \mathbf{X}(\ell,k) \\ &= Q_z(\ell,k) + Q_v(\ell,k), \end{aligned} \tag{10}$$

where $Q_z(\ell,k)$ and $Q_v(\ell,k)$ denote the residual reverberant and noise component at the beamformer's output. The spectral variance of $Q(\ell,k)$ is given by $\lambda_q(\ell,k) = E\{Q(\ell,k)(Q(\ell,k))^*\} = \lambda_{q_z}(\ell,k) + \lambda_{q_v}(\ell,k)$, where $(\cdot)^*$ denotes the complex conjugate, $\lambda_{q_z}(\ell,k)$ and $\lambda_{q_v}(\ell,k)$ denote the spectral variances of the residual reverberant and noise component at the beamformer's output. In addition, we can express $\lambda_{q_z}(\ell,k)$ as

$$\begin{aligned} \lambda_{q_z}(\ell,k) &= E\{Q_z(\ell,k)(Q_z(\ell,k))^*\} \\ &= \lambda_{q_\mathrm{e}}(\ell,k) + \lambda_{q_\ell}(\ell,k), \end{aligned} \tag{11}$$

where $\lambda_{q_\mathrm{e}}(\ell,k)$ and $\lambda_{q_\ell}(\ell,k)$ denote the residual early and late reverberation at the output of the beamformer. The spectral variance of the noise at the output of the MVDR beamformer is given by

$$\lambda_{q_v}(\ell,k) = \frac{1}{\mathbf{H}_\mathrm{d}^H(k)\mathbf{\Lambda}_{\mathbf{VV}}^{-1}(\ell,k)\mathbf{H}_\mathrm{d}(k)}. \quad (12)$$

3.2. Single-channel MMSE-LSA estimator

Assuming that the residual early and late reverberant signal components are mutually uncorrelated we can reduce the residual late reverberation at the output of the MVDR beamformer using a spectral enhancement technique. Here we employ a single-channel MMSE log spectral amplitude (LSA) estimator [7] to estimate the residual early speech component at the beamfomer's output.

Let $\xi(\ell,k)$ denote the *a priori* SIR,

$$\xi(\ell,k) = \frac{\lambda_{q_\mathrm{e}}(\ell,k)}{\lambda_{q_\ell}(\ell,k)+\lambda_{q_v}(\ell,k)}, \quad (13)$$

and $\gamma(\ell,k)$ denote the *a posteriori* SIR,

$$\gamma(\ell,k) = \frac{|Q(\ell,k)|^2}{\lambda_{q_\ell}(\ell,k)+\lambda_{q_v}(\ell,k)}. \quad (14)$$

The spectral variance $\lambda_{q_v}(\ell,k)$ of the residual noise is estimated using (12). In Section 4 we show how the LRSV $\lambda_{q_\ell}(\ell,k)$ is obtained.

Since $\lambda_{q_\mathrm{e}}(\ell,k)$ is unobservable we estimate the *a priori* SIR using the decision-directed approach [8]. The single-channel MMSE LSA gain function is given by [7]

$$G_\mathrm{LSA}(\ell,k) = \frac{\xi(\ell,k)}{1+\xi(\ell,k)} \exp\left(\frac{1}{2}\int_{\zeta(\ell,k)}^{\infty} \frac{e^{-t}}{t}\,\mathrm{d}t\right), \quad (15)$$

where

$$\zeta(\ell,k) = \frac{\xi(\ell,k)}{1+\xi(\ell,k)}\,\gamma(\ell,k). \quad (16)$$

To avoid speech distortions a lower bound, denoted by G_min, is applied to $G_\mathrm{LSA}(\ell,k)$. An estimate of the residual early spectral speech component $Q_\mathrm{e}(\ell,k)$ can now be obtained by applying the constraint gain function to the beamformer's output $Q(\ell,k)$, i.e.,

$$\hat{Q}_\mathrm{e}(\ell,k) = \max(G_\mathrm{LSA}(\ell,k), G_\mathrm{min})\,Q(\ell,k). \quad (17)$$

4. LATE REVERBERANT SPECTRAL VARIANCE ESTIMATORS

In this section we discuss two LRSV estimators. As illustrated in Fig. 1 the first estimator uses the output signal of the MVDR beamformer while the second estimator uses the received microphone signals. In the following we assume that the direct-to-reverberation ratio (DRR) is smaller than 0 dB.

4.1. Using the beamformer's output signal

Since the DRR is assumed to be smaller than 0 dB the LRSV is given by [1, 4, 5]

$$\lambda_{q_\ell}(\ell,k) = \mathrm{e}^{-2\alpha(k)RN_\mathrm{e}}\lambda_{q_z}(\ell-N_\mathrm{e},k). \quad (18)$$

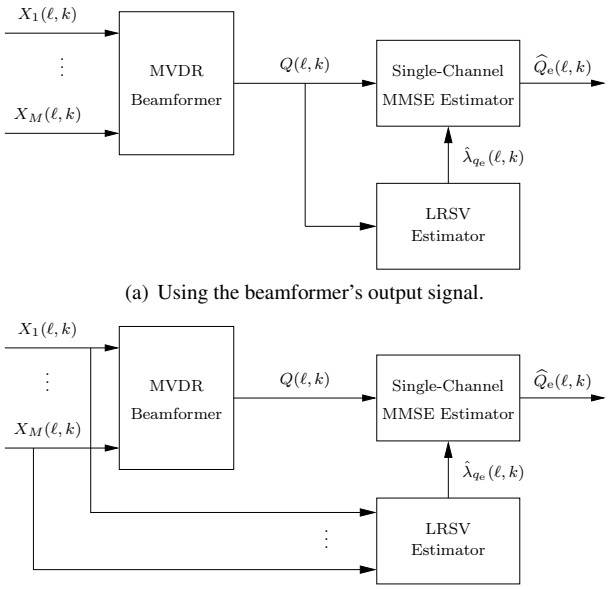
where R denotes the number of samples between two successive STFT frames, $\lambda_{q_z}(\ell,k)$ denotes the spectral variance of the residual reverberant component at the beamformer's output [as defined in (11)], and $\alpha(k)$ is linked to the frequency dependent reverberation time $T_{60}(k)$ through

$$\alpha(k) \triangleq \frac{3\,\log_\mathrm{e}(10)}{T_{60}(k)\,f_\mathrm{s}}. \quad (19)$$

In case $Q_v(\ell,k)=0$ the spectral variance of the reverberant component $\lambda_{q_z}(\ell,k)$ can be estimated by

$$\hat{\lambda}_{q_z}(\ell,k) = \eta\,\hat{\lambda}_{q_z}(\ell-1,k) + (1-\eta)\,|Q(\ell,k)|^2, \quad (20)$$

where η ($0 \le \eta < 1$) denotes the smoothing factor. In case $Q_v(\ell,k) \neq 0$ we need to estimate the spectral variance $\lambda_{q_z}(\ell,k)$ from $Q(\ell,k)$ before we can estimate the LRSV $\lambda_{q_\ell}(\ell,k)$. Details regarding this procedure can be found in [4, 5].



(a) Using the beamformer's output signal.

(b) Using the received microphone signals.

Fig. 1. Two approaches to estimate the LRSV.

4.2. Using the received microphone signals

Rather than using the output the MVDR beamformer we propose to exploit all microphone signals. In addition, we exploit *a priori* information about the spatial coherence of the reverberant sound field.

The LRSVs $\lambda_{z_{\ell,m}}(\ell,k)$ for $m \in \{1,\ldots,M\}$ are given by

$$\lambda_{z_{\ell,m}}(\ell,k) = \mathrm{e}^{-2\alpha(k)RN_\mathrm{e}}\lambda_{z,m}(\ell-N_\mathrm{e},k). \quad (21)$$

In case $V_m(\ell,k) \neq 0$ we first need to estimate the spectral variance $\lambda_{z,m}(\ell,k)$ before we can estimate the LRSV.

As shown in [4, 5] we can calculate the spatially averaged spectral variance using

$$\bar{\lambda}_{z_\ell}(\ell,k) = \frac{1}{M}\sum_{m=1}^{M}\lambda_{z_{\ell,m}}(\ell,k). \quad (22)$$

We now propose to calculate the spectral variance $\lambda_{q_\ell}(\ell,k)$ using

$$\lambda_{q_\ell}(\ell,k) = \bar{\lambda}_{z_\ell}(\ell,k)\mathbf{W}_\mathrm{MVDR}(\ell,k)^H\,\mathbf{\Gamma}(k)\,\mathbf{W}_\mathrm{MVDR}(\ell,k), \quad (23)$$

where $\mathbf{\Gamma}(k)$ denotes the spatial coherence matrix of the reverberant sound field. In case we assume that the reverberant sound field is spatially white we have

$$\mathbf{\Gamma}(k) = \mathbf{I}, \tag{24}$$

where $\mathbf{I}$ denotes the identity matrix of size $M \times M$. For $M = 2$ the resulting estimator is equivalent to the estimator used in [3]. A more realistic assumption is to assume that the reverberant sound field is spherically isotropic (i.e., diffuse). In the latter case the (p, q) element of the spatial coherence matrix is given by [9]

$$\gamma_{pq}(k) = \frac{\sin(\omega_k \, d_{pq}/c)}{\omega_k \, d_{pq}/c},$$

where c denotes the sound velocity and d_{pq} the distance between the pth and qth microphone.

4.3. Discussion

A major advantage of the single-microphone LRSV estimator is the low computational complexity. The computational complexity of the multi-microphone LRSV estimator is approximately a factor M higher.

Another advantage of the single-microphone LRSV estimator is the ability to deal with higher noise levels. Given an estimate of the covariance of the noise the MVDR beamformer is able to suppress noise. The spectral variance $\lambda_{q_v}(\ell, k)$ of the residual noise at the output of the MVDR beamformer can be calculated using (12) or estimated using the method proposed in [10]. The estimated spectral variance $\hat{\lambda}_{q_v}(\ell, k)$ can then be used to estimate $\lambda_{q_z}(\ell, k)$ as described in [4,5]. Depending on the noise field the power of the residual noise $\hat{\lambda}_{q_v}(\ell, k)$ can be significantly lower than the power of the noise at the received microphones. Therefore, it can be expected that the single-microphone LRSV estimator is able to cope with higher noise levels compared to the multi-microphone LRSV estimator.

A major advantage of the multi-microphone LRSV estimator is that we can exploit *a priori* knowledge of the reverberant sound field. Using statistical room acoustic theory it can be shown that the LRSV at each spatial position is equal. Therefore we observe different realizations of the same random process at different spatial positions. Hence, averaging the LRSV over different spatial positions significantly reduces the variance of the estimate. When the estimated LRSV is subsequently used by the MMSE estimator possible artifacts such as so-called musical tones (i.e., short duration random tones which can be very objectionable to the human ear) can be reduced.

5. EXPERIMENTAL RESULTS

In this section we evaluate the performance of different dereverberation algorithms. The algorithms were tested using reverberant speech (sampling rate was $f_s = 8$ kHz) that was generated by convolving anechoic speech fragments from the APLAWD database [11] with various AIRs. The AIRs were generated using an efficient implementation of the celebrated image method [12]. We used a uniform linear array with 5 microphones and an inter-microphone distance of 5 cm. The source was positioned at the broadside of the microphone array and the distance between the source and the center of the array was 3 m. The reverberation time T_{60} ranges from 200 to 800 ms. Here we assume that the noise field is spatially white such that the MVDR beamformer reduces to the well known delay and sum beamformer (DSB), the SNR was set to 30 dB.

Abbreviation	Description
SD	Single-microphone dereverberation applied to $x_1(n)$.
DSB	Delay and sum beamformer.
DSB + SD	DSB and single-microphone dereverberation.
DSB + $\mathrm{MD_A}$	DSB and multi-microphone dereverberation assuming a spatially white sound field.
DSB + $\mathrm{MD_B}$	DSB and multi-microphone dereverberation assuming a spherically isotropic sound field.

Table 1. Descriptions and abbreviations of the dereverberation algorithms.

The length of the STFT analysis and synthesis window was 256 samples, and an overlap between two successive STFT frames was 75% ($R = 64$). The parameter $G_{\min}$, which controls the maximum suppression, was set to -12 dB. The smoothing factor η was set to 0.9, and the weighting factor used in the decision-directed approach was set to 0.98. The time instance (measured with respect to the arrival time of the direct sound) at which the late reverberation starts was set to 48 ms, i.e., $N_e = 6$. The full-band reverberation time T_{60} was determined by applying Schroeder's method [13] to the AIR between the source and the first microphone. The decay rate $\alpha(k)$ was calculated using (19). In practice one can use a blind estimation procedure as proposed in [14, 15]. The first microphone was used as a reference for the time-alignment of the signals.

The performance of the algorithms was evaluated using the segmental signal to reverberation ratio (SRR) and log spectral distance (LSD) measures [4]. The direct sound signal received by the first microphone was used as a reference for these measures. For each reverberation time the results were averaged over 10 (male and female) speech fragments and 5 sentences. In addition, the results were averaged over 10 different source-array configurations that were obtained by translating and rotating the configuration in the enclosure such that the distances between the source and the microphones are invariant.

We evaluated one unprocessed microphone signal [$x_1(n)$] and the output signal of five different dereverberation algorithms, a description of the algorithms can be found in Table 1. The averaged results are shown in Figure 2. From these results it is clear that the first microphone signal (MIC) exhibits the lowest segmental SRR and highest LSD. Hence, compared to the first microphone signal all algorithms are able to increase the segmental SRR and lower the LSD. For reverberation times larger than 0.45 seconds the signal-microphone dereverberation algorithm (SD) outperforms the DSB that uses 5 microphones. The performance of the MMSE estimator that uses a LRSV estimate obtained from the output of the DSB beamformer (DSB+SD) provides better results compared to the DSB and single-channel MMSE estimator (SD). The performance is further increased when one of the proposed multi-microphone LRSV estimators (DSB + $\mathrm{MD_A}$ or DSB + $\mathrm{MD_B}$) is used. A marginal improvement is obtained under the assumption that the reverberant sound field is spatially white (DSB + $\mathrm{MD_A}$). However, a significant improvement is obtained under the assumption that the reverberant sound-field is spherically isotropic (DSB + $\mathrm{MD_B}$). Informal listening tests confirmed that late reverberation was reduced and indicated that the use of the multi-microphone LRSV estimator results in less musical tones and speech distortion compared to the single-microphone LRSV estimator.

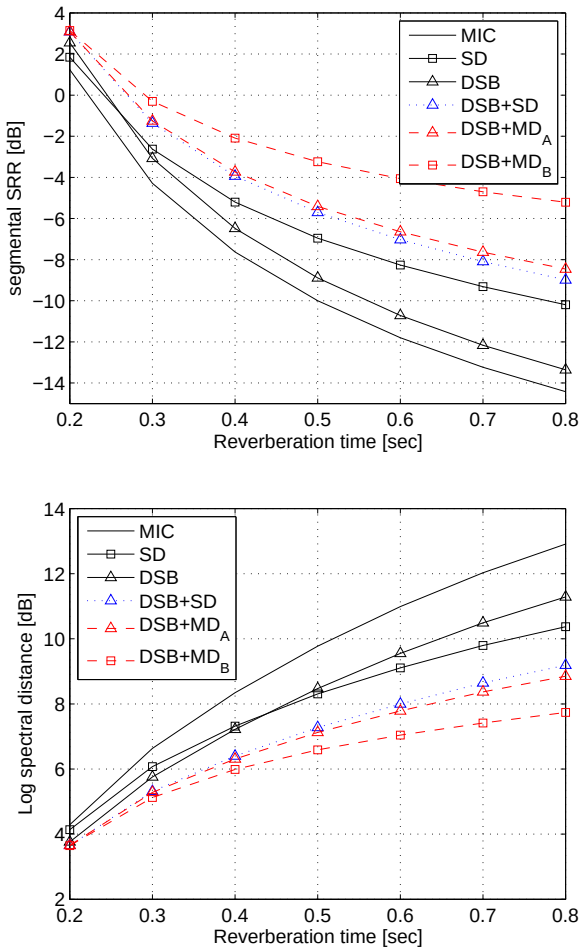


Fig. 2. Segmental SRRs and LSDs of all test signals. The reverberation time varies between 0.2 and 0.8 s ($M = 5$, $D = 3$ m, and $N_e R/f_s = 48$ ms).

6. CONCLUSIONS

In this paper a MVDR beamformer and a single-channel MMSE estimator were employed to reduce late reverberation. The single-channel MMSE estimator, which estimates the early speech component at the beamformer's output, requires an estimate of the LRSV. Here we proposed a LRSV estimator that utilizes all microphone signals and a statistical property of the reverberant sound field. The proposed multi-microphone LRSV estimator was compared with an existing LRSV estimator that utilizes only the output of the MVDR beamformer. We showed that when the proposed multi-microphone LRSV estimator is used by the MMSE estimator rather than the existing single-microphone LRSV estimator the performance in terms of segmental SRR and LSD is increased. A significant improvement was obtained under the assumption that the reverberant sound field is spherically isotropic. In addition, informal listening tests confirmed that the MMSE estimator that utilizes the LRSV estimate obtained by the proposed multi-microphone LRSV estimator introduces less musical tones and speech distortion compared to the single-microphone LRSV estimator.

7. REFERENCES

[1] K. Lebart, J.M. Boucher, and P.N Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoustica*, vol. 87, no. 3, pp. 359–366, 2001.

[2] E.A.P. Habets, "Multi-Channel Speech Dereverberation based on a Statistical Model of Late Reverberation," in *Proc. of the 30th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, Philadelphia, USA, Mar. 2005, pp. 173–176.

[3] E.A.P. Habets, S. Gannot, and I. Cohen, "Dual-microphone speech dereverberation in a noisy environment," in *Proc. of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT'06)*, Vancouver, Canada, Aug. 2006.

[4] E.A.P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.d. Thesis, Technische Universiteit Eindhoven, June 2007.

[5] E.A.P. Habets, "Speech dereverberation using spectral enhancement," in *Speech Dereverberation*, P.A. Naylor and N.D. Gaubitch, Eds. Springer, to appear in 2009.

[6] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. Mohan Sondhi, and Y. Huang, Eds., chapter 48. Springer, 2007, part H.

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[8] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[9] B.F. Cron and C.H. Sherman, "Spatial-correlation functions for various noise models," *Journal of the Acoustical Society of America*, vol. 34, no. 11, pp. 1732–1736, Nov. 1962.

[10] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 5, pp. 466–475, Sept. 2003.

[11] G. Lindsey, A. Breen, and S. Nevard, "SPAR's archivable actual-word databases," Tech. Rep., University College London, June 1987.

[12] E. A. P. Habets, "Room impulse response (RIR) generator," [Online] Available: http://home.tiscali.nl/ehabets/rir_generator.html, Oct. 2008.

[13] M. R. Schroeder, "A new method of measuring reverberation time," *Journal of the Acoustical Society of America*, vol. 37, pp. 409–412, 1965.

[14] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien, Jr., C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," *Journal of the Acoustical Society of America*, vol. 114, no. 5, pp. 2877–2892, Nov. 2003.

[15] H. W. Löllmann and P. Vary, "Estimation of the reverberation time in noisy environments," in *International Workshop on Acoustic Echo and Noise Control (IWAENC'08)*, Sep 2008, pp. 1–4.