NESTED GENERALIZED SIDELOBE CANCELLER FOR JOINT DEREVERBERATION AND NOISE REDUCTION

Ofer Schwartz[1]*, Sharon Gannot*[1]*, and Emanuël A. P. Habets*[2]

[1]Faculty of Engineering, Bar-Ilan University, Ramat-Gan, 5290002, Israel
[2]International Audio Laboratories Erlangen*, Am Wolfsmantel 33, 91058 Erlangen, Germany

ABSTRACT

Speech signal is often contaminated by both room reverberation and ambient noise. In this contribution, we propose a nested generalized sidelobe canceller (GSC) beamforming structure, comprising an *inner* and an *outer* GSC beamformers (BFs), that decouple the speech dereverberation and the noise reduction operations. The BFs are implemented in the short-time Fourier transform (STFT) domain. Two alternative reverberation models are adopted. In the first, used in the inner GSC, reverberation is assumed to comprise a coherent early component and a late reverberant component. In the second, used in the outer GSC, the influence of the entire acoustic transfer function (ATF) is modeled as a convolution along the frame index in each frequency. Unlike other BF designs for this problem that must be updated in each time-frame, the proposed BF is time-invariant in static scenarios. Experiments with both simulated and recorded environments verify the effectiveness of the proposed structure.

Index Terms— Dereverberation, Noise reduction.

1. INTRODUCTION

Reverberant and noisy speech can be difficult to understand for both humans and machines, and can lead to listening fatigue. Methods that are able to reduce both reverberation and noise therefore play an important role in, for example, speech communication systems and hearing aids. While many solutions focus only dereverberation (c.f. [1] and the references therein) or noise reduction, only few focus on both noise and reverberation reduction [2–6].

The minimum variance distortionless response (MVDR) BF, usually implemented in a GSC structure [7], is a popular noise reduction algorithm [8, 9]. The GSC consists of two branches, namely, an upper branch that is responsible for maintaining a desired response towards the signal of interest, and a lower branch that is responsible for interference reduction. The two branches are usually orthogonal. In earlier works, the relative transfer functions (RTFs) that describe the inter-channel relations were modeled using multiplicative transfer function (MTF). In [10] the RTFs were modelled as a convolutive transfer function (CTF) to handle higher reverberation levels while using short processing frames. Similar to [9], the algorithm in [10] focuses on noise reduction and yields a reverberant output signal. In [11] the fixed beamformer (FBF) in the upper branch of the GSC is replaced by a delay and sum beamformer (DSBF), while the design of the blocks in the lower branch remain unaltered. The DSBF, which is known to maximize the white noise gain (WNG), is also able to reduce some early reflections and late reverberation. It is interesting to note that the branches of the resulting GSC are non-orthogonal.

This research was partially supported by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development.

*A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

In [12], a structure comprised of the MVDR BF and a corresponding postfilter was proposed. The MVDR was designed to suppress noise and early reflections while maintaining the direct path. The postfilter was designed to suppress the residual noise and late reverberation. The late reverberant field was assumed to be an ideal diffuse sound field, and the late reverberation level was estimated using Polack's model [13]. Recently, in [14], a multichannel Wiener filter (MWF) BF, decomposed into an MVDR BF and subsequent postfilter, was proposed. The MVDR component suppresses both the late reverberation and the noise, while the early reflections are reduced by using a DSBF in the upper branch similarly to [11]. Since the noise plus late reverberant spatial correlation matrix is time-varying, the noise canceller (NC) in the lower branch needs to be recomputed for each time-frequency bin.

In the current paper, a nested GSC approach, which aims to reduce both noise and reverberation, is proposed. The main idea of the proposed approach is the decoupling of the dereverberation and the noise reduction operations. The proposed structure consists of an *outer* GSC and an *inner* GSC. In the outer GSC the blocking matrix (BM) is designed to entirely block the reverberant signal. The outer GSC will therefore be only responsible of noise reduction. The FBF of the outer GSC is implemented using a second GSC, which is referred to as the inner GSC, that aims at reducing the late reverberation. In static scenarios with stationary noise, both the inner and outer GSCs can be implemented using time-invariant filters, which significantly reduces the computational complexity.

2. PROBLEM FORMULATION

The problem is formulated in the STFT domain with m denoting the time index and k denoting the frequency index. The nth microphone signal is given by

$$Y_n(m,k) = X_n(m,k) + V_n(m,k),\ n = 1,\dots,N \qquad (1)$$

where $X_n(m,k)$ denotes the reverberant speech, $V_n(m,k)$ denotes additive noise, and N is the number of microphones.

When the reverberation time is larger than the frame length,, the reverberant speech signal $X_n(m,k)$ cannot be efficiently modelled as a multiplication in the frequency domain, i.e., the MTF model is not valid [10]. We therefore propose two alternative models.

According to the first model, the reverberant speech as received by the n-th microphone, $X_n(m,k)$, is expressed as

$$X_n(m,k) = X_{\mathrm{e},n}(m,k) + X_{\ell,n}(m,k), \qquad (2)$$

where $X_{\mathrm{e},n}(m,k)$ denotes the early speech component that includes the direct path and early reflections, and $X_{\ell,n}(m,k)$ denotes the late reverberation. In this model, the early speech component is modelled using the MTF approximation such that

$$X_{\mathrm{e},n}(m,k) = G_{\mathrm{e},n}(k) \cdot X_{\mathrm{e},1}(m,k) \qquad (3)$$

where $G_{\mathrm{e},n}(k)$ denotes the relative early transfer function (RETF). This model can be summarized in a vector form as

$$\mathbf{y}(m,k) = \mathbf{g}_{\mathrm{e}}(k)\, X_{\mathrm{e},1}(m,k) + \mathbf{x}_{\ell}(m,k) + \mathbf{v}(m,k). \quad (4)$$

According to the second model, the reverberant speech $X_n(m,k)$ is expressed as

$$X_n(m,k) = \sum_{\widetilde{m}=-q_1}^{q_2} G_n(\widetilde{m},k) \cdot X_1(m+\widetilde{m},k), \quad (5)$$

where $G_n(m,k)$ are the relative CTF coefficients, and q_1, q_2 are the lengths of the causal and the noncausal parts of the relative CTF, respectively.

Following the idea presented in [14], our goal is to obtain an estimate of a spatially filtered version of the early speech components that is given by[1]

$$S_F(m) = \mathbf{h}_{\mathrm{DS}}^H\, \mathbf{g}_{\mathrm{e}}\, X_{\mathrm{e},1}(m), \quad (6)$$

where $\mathbf{h}_{\mathrm{DS}}$ is a DSBF steered towards the desired source and hence captures a spatially filtered version of the early speech component, which in the absence of steering errors includes the direct-path signal and spatially filtered early reflections.

In [14], an optimal MTF based MVDR BF followed by a post-filter was proposed to jointly reduce reverberation and noise. The output of the BF provided an estimate of $S_F(m)$ using

$$\widehat{S}_F(m) = \mathbf{h}_{\mathrm{MVDR}}^H(m)\, \mathbf{y}(m), \quad (7)$$

where

$$\mathbf{h}_{\mathrm{MVDR}}(m) = \underset{\mathbf{h}}{\operatorname{argmin}}\, \mathbf{h}^H \mathbf{\Phi}(m) \mathbf{h} \quad \text{s.t.} \quad \mathbf{h}^H \mathbf{g}_{\mathrm{e}} = \mathbf{h}_{\mathrm{DS}}^H \mathbf{g}_{\mathrm{e}} \quad (8)$$

where $\mathbf{\Phi}(m) = \mathbf{\Phi}_v(m) + \mathbf{\Phi}_{\ell}(m)$ denotes the power spectral density (PSD) matrix of the total interference where

$$\mathbf{\Phi}_v(m) = E\{\mathbf{v}(m)\mathbf{v}^H(m)\} \;\text{ and }\; \mathbf{\Phi}_{\ell}(m) = E\{\mathbf{x}_{\ell}(m)\mathbf{x}_{\ell}^H(m)\}$$

denote the PSDs matrices of the noise and late reverberation.

When we model the late reverberant sound field as an isotropic and homogeneous diffuse sound field, with a frequency dependent spatial coherence matrix $\mathbf{\Gamma}(k)$, the PSD matrix $\mathbf{\Phi}_{\ell}(m)$ can be written as

$$\mathbf{\Phi}_{\ell}(m) = \phi_{\ell}(m)\, \mathbf{\Gamma}, \quad (9)$$

where $\phi_{\ell}(m,k)$ is the PSD of the late reverberation.

In [14], the MVDR was implemented using a single non-orthogonal GSC. One of the advantages of the non-orthogonal GSC is that the FBF remains very simple. The computation of the filter coefficients of the noise-plus-reverberation canceller however requires an estimate of the noise PSD matrix and the late reverberation power $\phi_{\ell}(m)$. Moreover, since $\phi_{\ell}(m)$ is time-varying, the interference PSD matrix $\mathbf{\Phi}(m)$ should be recomputed for each time instant. As the inverse of the interference PSD matrix is also required, we also needed to recompute the inverse for each time instant.

Extending the criterion in (8) to the CTF model is a cumbersome task, since an equivalent CTF model to (9) that takes into account the inter-frame correlations is currently unavailable. Therefore, we propose a new nested GSC structure that decouples the dereverberation and noise reduction tasks, utilizing the two alternative signal models. It should be noted that the solution obtained using the proposed structure is not equal to the one obtained by (8).

[1]For brevity we omit the frequency index in the sequel when possible.

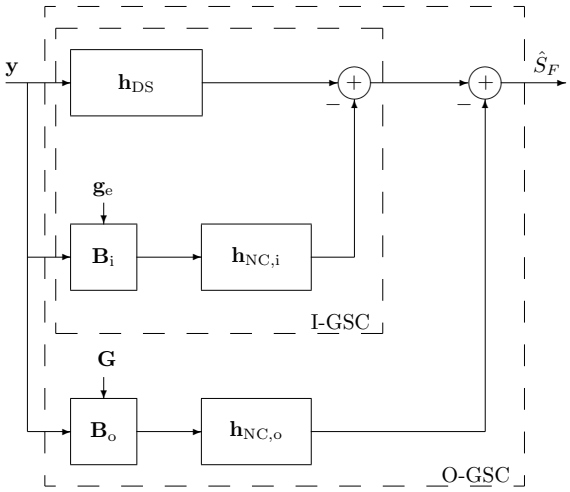

Fig. 1: Block diagram of the proposed nested GSC structure.

3. NESTED GENERALIZED SIDELOBE CANCELLER

We now propose a new structure, based on the GSC formulation, that decouples the noise and reverberation reduction processes, utilizing the two signal models described in Sec. 2. Specifically, a special structure that consists of two nested GSCs is proposed. The *inner* GSC that uses the model in (2), is solely responsible for dereverberation while the *outer* GSC that uses the model in (5), is solely responsible for the noise reduction. The proposed nested GSC structure is illustrated in Fig. 1.

3.1. Dereverberation using the inner GSC

The inner MVDR BF $\mathbf{h}_{\mathrm{MVDR,i}}$ (see Fig. 1) is designed to only reduce reverberation and therefore ignores the noise component.

It can be deduced from the block diagram in Fig. 1, that the inner FBF should satisfy the constraint of the entire system (8), i.e., $\mathbf{h}_{\mathrm{MVDR,i}}^H \mathbf{g}_e = \mathbf{h}_{\mathrm{DS}}^H \mathbf{g}_{\mathrm{e}} = F$. An MVDR BF that minimizes the late reverberation subject to this constraint is given by

$$\mathbf{h}_{\mathrm{MVDR,i}} = \underset{\mathbf{h}}{\operatorname{argmin}}\, \mathbf{h}^H \mathbf{\Phi}_{\ell}(m) \mathbf{h} \quad \text{s.t.} \quad \mathbf{h}^H \mathbf{g}_e = F. \quad (10)$$

The MVDR BF is implemented in a GSC structure [7, 9] (see [14] for a discussion on the non-orthogonal GSC implementation and its advantages):

$$\mathbf{h}_{\mathrm{MVDR,i}} = \mathbf{h}_{\mathrm{q,i}} - \mathbf{B}_i\, \mathbf{h}_{\mathrm{NC,i}}. \quad (11)$$

where $\mathbf{h}_{\mathrm{q,i}} = \mathbf{h}_{\mathrm{DS}}$ is a FBF such that the constraint $\mathbf{h}_{\mathrm{q,i}}^H \mathbf{g}_e = F$ is satisfied, $\mathbf{B}_{\mathrm{i}}$ is a BM responsible for blocking the early speech component, i.e., $\mathbf{B}_{\mathrm{i}}^H \mathbf{g}_e = \mathbf{0}$, and $\mathbf{h}_{\mathrm{NC,i}}$ is the respective NC, responsible for cancelling the residual reverberation at the FBF output.

The BM of the inner GSC is given by [9]:

$$\mathbf{B}_i = \begin{bmatrix} -G_{e,2}^* & -G_{e,3}^* & \dots & -G_{e,N}^* \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}. \quad (12)$$

The NC of the inner MVDR is implemented in closed-form due to the strong non-stationarity of the interference signal (late reverberation in this case):

$$\mathbf{h}_{\mathrm{NC,i}} = (\mathbf{B}_i^H\, \mathbf{\Gamma}\, \mathbf{B}_i)^{-1} \mathbf{B}_i^H\, \mathbf{\Gamma}\, \mathbf{h}_{\mathrm{DS}}. \quad (13)$$

Note that (22) is independent of the PSD $\phi_\ell(m)$, since it is cancelled out. Hence, for a static microphone constellation $\mathbf{h}_{\mathrm{NC,i}}$ does not have to be recomputed for each time instant, resulting in a significant reduction in the computational complexity. It should be stressed that a reduction in the noise level at the output of the inner GSC cannot be guaranteed.

3.2. Noise reduction using the outer GSC

The outer MVDR BF, $\mathbf{h}_{\mathrm{MVDR,o}}(m)$, is responsible for minimizing the noise component at the output of the inner GSC, thereby ignoring the reverberation. To avoid reverberation amplification, the entire reverberant signal should be blocked by the BM of the outer GSC. This can be obtained by adopting the CTF model [10] to describe the reverberation tail. While the original contribution [10] presented batch processing of the entire dataset, we adopt here a sequential representation as follows. Since the application of the GSC under the CTF model necessitates an extended definition of the involved signals, to incorporate past and future observations, we define a state-vector comprised of concatenated frames as

$$\widetilde{\mathbf{y}}_n(m) = \left[\begin{array}{cccc} Y_n(m-z_1) & \cdots & Y_n(m+z_2) & \end{array}\right]^T,\ n=2,\ldots,M$$
$$\bar{\mathbf{y}}_1(m) = \left[\begin{array}{cccc} Y_1(m-z_1-q_1) & \cdots & Y_1(m+z_2+q_2) & \end{array}\right]^T.$$

The reason for the different definition of the first microphone state-vector is explained later. The auxiliary observation vector can be written as

$$\widetilde{\mathbf{y}}_n(m) = \mathbf{G}_n\,\bar{\mathbf{x}}_1(m) + \widetilde{\mathbf{v}}_n(m),\ n=1,2,\ldots,M \tag{14}$$

where $\mathbf{G}_n,\ n=2,\ldots,M$ is the corresponding convolution matrix:

$$\mathbf{G}_n = \tag{15}$$
$$\left[\begin{array}{cccccc} G_n(-q_1) & \cdots & G_n(q_2) & 0 & 0 & \cdots \\ 0 & G_n(-q_1) & \cdots & G_n(q_2) & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & G_n(-q_1) & \cdots & G_n(q_2) \end{array}\right]. \tag{16}$$

In addition, $\mathbf{G}_1$ is defined as an identity matrix with proper dimensions, and

$$\bar{\mathbf{x}}_1(m) = \left[\begin{array}{cccc} X_1(m-z_1-q_1) & \cdots & X_1(m+z_2+q_2) & \end{array}\right]^T$$
$$\widetilde{\mathbf{v}}_i(m) = \left[\begin{array}{cccc} V_i(m-z_1) & \cdots & V_i(m+z_2) & \end{array}\right]^T.$$

Concatenating the N signal vectors yields

$$\widetilde{\mathbf{y}}(m) = \mathbf{G}\,\bar{\mathbf{x}}_1(m) + \widetilde{\mathbf{v}}(m), \tag{17}$$

where

$$\widetilde{\mathbf{y}}(m) = \left[\begin{array}{cccc} \bar{\mathbf{y}}_1^T(m) & \widetilde{\mathbf{y}}_2^T(m) & \ldots & \widetilde{\mathbf{y}}_N^T(m) \end{array}\right]^T$$
$$\mathbf{G} = \left[\begin{array}{cccc} \mathbf{G}_1^T & \mathbf{G}_2^T & \ldots & \mathbf{G}_N^T \end{array}\right]^T$$
$$\widetilde{\mathbf{v}}(m) = \left[\begin{array}{cccc} \bar{\mathbf{v}}_1^T(m) & \widetilde{\mathbf{v}}_2^T(m) & \ldots & \widetilde{\mathbf{v}}_N^T(m) \end{array}\right]^T.$$

The excess length of $\bar{\mathbf{y}}_1(m)$ is necessary such that all frames in $\bar{\mathbf{x}}_1(m)$ will be available for blocking the desired speech at the output of the outer BM. $\bar{\mathbf{v}}1(m)$ is defined similarly to $\bar{\mathbf{x}}_1(m)$.

The CTF-MVDR criterion for the outer BF can now be stated:

$$\mathbf{h}_{\mathrm{MVDR,o}} = \underset{\mathbf{h}}{\operatorname{argmin}}\,\mathbf{h}^H\widetilde{\mathbf{\Phi}}_v\mathbf{h} \quad \text{s.t.}\quad \mathbf{h}^H\mathbf{G} = \mathbf{f}^H\mathbf{G} \tag{18}$$

where $\widetilde{\mathbf{\Phi}}_v = E\{\widetilde{\mathbf{v}}(m)\widetilde{\mathbf{v}}^H(m)\}$ and $\mathbf{f}$ is given by

$$\mathbf{f}^T = \left[\begin{array}{ccccccc} \mathbf{0}_{z_1+q_1}^T & H_{1,\mathrm{MVDR,i}} & \mathbf{0}_{z_2+q_2}^T & & & & \\ \mathbf{0}_{z_1}^T & H_{2,\mathrm{MVDR,i}} & \mathbf{0}_{z_2}^T & \ldots & \mathbf{0}_{z_1}^T & H_{M,\mathrm{MVDR,i}} & \mathbf{0}_{z_2}^T \end{array}\right], \tag{19}$$

where $H_{n,\mathrm{MVDR,i}},\ n=1,\ldots,N$ are the elements of $\mathbf{h}_{\mathrm{MVDR,i}}$ and $\mathbf{0}_\alpha$ is a vector of α zeros, necessary to constrain only the current frames in accordance with (8).

The outer MVDR $\mathbf{h}_{\mathrm{MVDR,o}}$ can be efficiently decomposed and implemented as a GSC such that

$$\mathbf{h}_{\mathrm{MVDR,o}} = \mathbf{h}_{\mathrm{q,o}} - \mathbf{B}_{\mathrm{o}}\mathbf{h}_{\mathrm{NC,o}} \tag{20}$$

with $\mathbf{h}_{\mathrm{q,o}} = \mathbf{f}$, satisfying, due to definition (19), $\mathbf{h}_{\mathrm{q,o}}^H\widetilde{\mathbf{y}}(m) = \mathbf{h}_{\mathrm{MVDR,i}}^H\mathbf{y}(m)$. The BM is given by

$$\mathbf{B}_{\mathrm{o}} = \left[\begin{array}{cccc} -\mathbf{G}_2^H & -\mathbf{G}_3^H & \cdots & -\mathbf{G}_N^H \\ \mathbf{I} & 0 & \cdots & 0 \\ 0 & \mathbf{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \mathbf{I} \end{array}\right]. \tag{21}$$

The NC $\mathbf{h}_{\mathrm{NC,o}}(m)$ is responsible for mitigating the residual noise at the output of the outer FBF $\mathbf{h}_{\mathrm{q,o}}$, which is equivalent in our case to the inner MVDR $\mathbf{h}_{\mathrm{MVDR,i}}$. A closed-form solution is given by

$$\mathbf{h}_{\mathrm{NC,o}} = (\mathbf{B}_{\mathrm{o}}^H\,\widetilde{\mathbf{\Phi}}_v, \mathbf{B}_{\mathrm{o}})^{-1}\mathbf{B}_{\mathrm{o}}^H\,\widetilde{\mathbf{\Phi}}_v\,\mathbf{f}. \tag{22}$$

Note, that in static scenarios $\widetilde{\mathbf{\Phi}}_v$ and $\mathbf{B}_{\mathrm{o}}$ are time-invariant, and hence the matrix inversion needs to be applied only once.

3.3. Parameters Estimation

The nested GSC requires the estimation of two sets of parameters: 1) the coherence matrix of the late reverberation and the noise PSD matrix $\widetilde{\mathbf{\Phi}}_v$, 2) the relative CTF coefficients $G_n(m)$ and the RETF $G_{e,n}$.

By modelling the late reverberation as a spherically isotropic and homogeneous diffuse noise field, the (i,j)th element of $\mathbf{\Gamma}$ is given by $\mathrm{sinc}\left(\frac{f_s k d_{1,1}}{Kc}\right)$ with K the number of frequency bins, $d_{i,j}$ the inter-distance between microphones i and j, c the sound velocity, and f_s the sampling frequency.

The noise PSD matrices can be estimated during speech-absence, by using an estimate of the speech presence probability (c.f. [15]). In this contribution the availability of speech-absent frames is assumed.

The relative CTF can be estimated by utilizing the speech non-stationarity [16]. For estimating the RETF we propose to first obtain an estimate of the early component at each microphone $\widehat{X}_{\mathrm{e},n}(m)$, utilizing single-channel dereverberation technique (c.f. [17]). Then, using least squares (LS) fit between the early components of a microphone pair, $G_{\mathrm{e},n}(m),\ n=1,\ldots,N$ can be identified. More details can be found in [14].

4. PERFORMANCE EVALUATION

The performance of the proposed algorithm is evaluated in terms of two objective measures that are commonly used in the speech enhancement community, namely perceptual evaluation of speech quality (PESQ) [18] and log spectral distance (LSD). The PESQ

scores and LSD measure were computed by averaging the results obtained using 82 sentences 12–24 Sec long, evenly distributed between female and male speakers. The CTF and the RETF coefficients were re-estimated for each sentence. The coefficients were blindly estimated from the noisy and reverberant sentence using the LS technique as described above. The clean reference for evaluation in all cases was the anechoic speech $S(m)$ filtered by the average of the early transfer functions $1/N \sum_{n=1}^{N} G_{\mathrm{e},n}$ (note, that since in our case the microphone signals are aligned, the DSBF simplifies to an average). Two scenarios were considered: 1) simulated reverberant signals plus directional noise; 2) recorded reverberant signals contaminated by recorded air-condition noise. In both cases sensor noise was added to the microphone signals with 50 dB signal to noise ratio (SNR).

The performance of the proposed algorithm was compared with six competing algorithms: 1) DSBF; 2) The MVDR BF in [14] without postfilter, as described in (8); 3) The inner GSC BF; 4) The outer GSC BF with the lower branch implemented with the MTF model; 5) The outer GSC BF with the lower branch implemented with the CTF model (similarly to [11]); 6) The nested GSC with $z_1 = z_2 = q_1 = q_2 = 0$ (i.e., assuming that the MTF model also holds for the outer GSC). The desired signal component at the output of all algorithms is constrained to be equal to the output of a DSBF of the early components, to guarantee fair comparison between them.

4.1. Simulated Data

The room impulse responses (RIRs) were computed using an efficient implementation of the image method [19, 20]. Room dimensions were set to $[6.1,\ 5.3,\ 2.7]$ m and the reverberation time was set to $T_{60} = 0.6$ s. The desired source was located at the broadside of the array and the noise source at the endfire. The average distance of both sources was set to 3 m. Sampling rate was set to 16 kHz. Directional noise was added to the simulated reverberant signals to obtain various SNR levels. The parameters of the CTF model in the outer GSC were $q_1 = 8, q_2 = 2, z_1 = 2, z_2 = 1$. A diagonal loading with value of 10^{-4} was added to $\mathbf{\Gamma}(m)$. The noise PSD, $\widetilde{\mathbf{\Phi}}_v(m)$, was estimated from the microphones during speech absence (assuming an ideal voice activity detector). In Table 1 the results for several SNR levels are depicted. The proposed multichannel algorithm clearly outperforms the competing algorithms.

4.2. Recorded Data

Reverberant signals and air-conditioning noise were (separately) recorded in the var-echoic acoustic laboratory at Bar-Ilan University, Israel. The speech utterances were played in the room using a Fostex 6301BX loudspeaker and were recorded by four AKG CK32 omnidirectional microphones, mounted on a metal ruler. The room dimensions are $[6,\ 6,\ 2.4]$ m. Reverberation time was set by adjusting the room panels, and was measured to be approximately $T_{60} = 0.5$ s. The reverberant speech and the air-condition noise signals were mixed in several SNR levels. The spatial PSD matrix $\widetilde{\mathbf{\Phi}}_v(m)$, was estimated using periods in which the desired speech source was inactive. The parameters of the CTF model in the outer GSC were $q_1 = 6, q_2 = 1, z_1 = 4, z_2 = 1$. A diagonal loading with value of 10^{-2} was added to the diffused coherence matrix $\mathbf{\Gamma}(m)$. In Table 2 the results for several SNR levels are depicted. The proposed algorithms outperforms all competing algorithm w.r.t. the LSD measure. In PESQ scores, the proposed algorithm achieves better or equivalent results compared with the competing algorithms.

LSD	5 dB	10 dB	15 dB	20 dB
Unprocessed	6.56	5.83	5.21	4.73
DSBF	4.87	4.40	4.02	3.74
MVDR w. MTF (8)	3.93	3.58	3.33	3.15
Inner GSC	4.62	4.16	3.81	3.55
Outer GSC w. MTF	3.92	3.57	3.32	3.15
Outer GSC w. CTF [11]	3.93	3.55	3.30	3.14
Nested w. MTF	3.88	3.54	3.30	3.14
Nested w. CTF	**3.71**	**3.37**	**3.15**	**3.01**
PESQ	5 dB	10 dB	15 dB	20 dB
Unprocessed	1.08	1.12	1.18	1.25
DSBF	1.11	1.15	1.22	1.28
MVDR w. MTF (8)	1.20	1.28	1.36	1.43
Inner GSC	1.12	1.18	1.25	1.33
Outer GSC w. MTF	1.21	1.28	1.35	1.42
Outer GSC w. CTF [11]	1.27	1.35	1.41	1.46
Nested w. MTF	1.21	1.28	1.36	1.42
Nested w. CTF	**1.29**	**1.38**	**1.45**	**1.50**

Table 1: Simulated reverberant signals plus spatially-white noise for a speaker-array distance of 3 m.

LSD	5 dB	10 dB	15 dB	20 dB
Unprocessed	5.60	5.02	4.54	4.17
DSBF	4.42	4.04	3.73	3.5
MVDR w. MTF (8)	3.85	3.53	3.29	3.11
Inner GSC	4.22	3.85	3.56	3.33
Outer GSC w. MTF	3.82	3.51	3.28	3.10
Outer GSC w. CTF [11]	3.79	3.49	3.29	3.10
Nested w. MTF	3.81	3.51	3.27	3.10
Nested w. CTF	**3.73**	**3.42**	**3.19**	**3.01**
PESQ	5 dB	10 dB	15 dB	20 dB
Unprocessed	1.10	1.17	1.26	1.36
DSBF	1.15	1.23	1.33	1.44
MVDR w. MTF (8)	1.24	1.35	1.46	1.57
Inner GSC	1.16	1.25	1.35	1.47
Outer GSC w. MTF	1.26	1.36	1.46	1.57
Outer GSC w. CTF [11]	**1.27**	1.37	1.47	1.57
Nested w. MTF	1.26	1.36	1.46	1.57
Nested w. CTF	**1.27**	**1.38**	**1.49**	**1.60**

Table 2: Recorded reverberant signals for at a source-array distance of 3 m plus air-condition noise.

5. CONCLUSIONS

A nested GSC structure to decouple late reverberation reduction and noise reduction is proposed. The proposed structure is comprised of an inner GSC, which suppresses the late reverberation while preserving the early speech components, and an outer GSC which reduces the noise. The BM of the outer GSC is implemented using the CTF model of the acoustic path to guarantee proper blocking of the entire reverberant tail. An important attribute of the proposed structure is its time-invariance. Optimality of the nested GSC is not claimed. The experimental study consists of both simulated data and recordings in actual acoustic environment, demonstrating the performance advantages of the proposed structure over several competing algorithms.

6. REFERENCES

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

[2] H. Attias and L. Deng, "Speech denoising and dereverberation using probabilistic models," *Advances in Neural Information Processing Systems (NIPS)*, vol. 13, pp. 758–764, 2001.

[3] S. Doclo and M. Moonen, "Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Darmstadt, Germany, Sept. 2001, pp. 31–34.

[4] E. A. P. Habets, *Single- and multi-microphone speech dereverberation using spectral enhancement*, Ph.D. Thesis, Technische Universiteit Eindhoven, June 2007.

[5] H. W. Löllmann and P. Vary, "Low delay noise reduction and dereverberation for hearing aids," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1–9, Jan. 2009.

[6] E. A. P Habets and J. Benesty, "A two-stage beamforming approach for noise reduction and dereverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 945–958, May 2013.

[7] L. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, 1982.

[8] S. Affès and Y Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 425–437, 1997.

[9] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, 2001.

[10] R. Talmon, I. Cohen, and S. Gannot, "Convolutive transfer function generalized sidelobe canceler," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 7, pp. 1420–1434, Sep. 2009.

[11] R. Talmon, I. Cohen, and S. Gannot, "Multichannel speech enhancement using convolutive transfer function approximation in reverberant environments," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3885–3888.

[12] E. A. P. Habets, "Towards multi-microphone speech dereverberation using spectral enhancement and statistical reverberation models," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, 2008, pp. 806–810.

[13] J. D. Polack, *La transmission de l'énergie sonore dans les salles*, Ph.D. thesis, Université du Maine, Le Mans, France, 1988.

[14] O. Schwartz, S. Gannot, and E. A.P. Habets, "Multi-microphone speech dereverberation and noise reduction using relative early transfer functions," submitted to IEEE Trans. on Audio, Speech, and Lang. Process., Mar. 2014.

[15] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, 2012.

[16] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 4, pp. 546–555, May 2009.

[17] E. A. P. Habets, "Speech dereverberation using statistical reverberation models," in *Speech Dereverberation*, P. A. Naylor and N. D. Gaubitch, Eds. Springer, 2010.

[18] ITU-T, "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.

[19] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. of Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[20] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. of Acoustical Society of America*, vol. 76, no. 5, pp. 1527–1529, Nov. 1986.