

# Adaptive Beamforming in Room with Reverberation

Zoran Šarić<sup>1</sup> and Slobodan Jovičić<sup>2</sup>

<sup>1</sup>Institute of Security, Belgrade, Serbia and Montenegro

<sup>2</sup>Faculty of Electrical Engineering, University of Belgrade, Serbia and Montenegro

e-mail: sare@yubc.net

## Abstract

Microphone arrays are powerful tools for noise suppression in a reverberant room. Generalized Sidelobe Canceller (GSC) that exploits Minimum Variance (MV) criterion is efficient in interference suppression when there is no correlation between the desired signal and the interferences. Correlation between the desired signal and any of interferences produces a desired signal cancellation and degradation of signal-to-noise ratio. This paper analyses the unwanted cancellation of the desired source. It shows that cancellation level of the desired signal is proportional to the correlation between the direct wave and the reflected waves. For prevention of a desired signal cancellation we suggest the GSC parameter estimation during the pauses of the desired signal. For this case it is analytically shown that there is no cancellation of the desired signal. The proposed algorithm was experimentally tested and compared with the Conventional Beamformer (CBF) and GSC. Experimental tests have shown the advantage of the proposed method.

## 1. Introduction

Microphone array is a powerful tool for the noise suppression in the room with reverberation. It acts as a spatial selective filter which enhances the desired signal and attenuates interferences. Properties of a microphone array depend on the algorithm used to adjust its weightings. The simplest method is the *Conventional Beamformer* (CBF). This method has fixed weightings that form the desired directivity pattern. CBF is optimal when microphone array is in diffuse a noise field.

When microphone array is in a diffuse noise field, Claude Marro at al. [1] propose the postfiltering algorithm based on nulls in coherence function. This array has fixed non adaptive weightings and post processing with the Wiener filter.

The adaptive beamformer can be realized as Frost's beamformer or GSC (*Generalized Sidelobe Canceller*). Both structures are completely equivalent. They are commonly referred to as *Minimum Variance Distortionless Response* (MVDR) beamformer. These algorithms adapt their beam pattern so as to maximally suppress interferences having unity gain for the desired signal [2],[3].

In [4] partially adaptive sidelobe canceller is proposed that exploits a prior knowledge of the positions of some interferences. The rest of interferences whose positions are unknown, are suppressed adaptively. Like GSC this algorithm has good performance in the coherent signal space.

MVDR beamformers are sensitive to leakage of desired signal into reference signal space. This leakage is present in the room with reverberation where there are lot of reflections both desired signal and interferences. The consequence is an

unwanted cancellation of the desired source and degradation of the signal to noise (SNR) improvement.

MV criterion can not resolve the problem of cancellation of the desired signal. One of the possible solutions is to replace MV criterion with minimization of cross-correlation between outputs [5]. The method in [5] exploits geometric information about source positions which is often used in adaptive beamforming, and combines it with blind source separation criteria such as signal independence. In the practice the locations of the sources are not a priori known. To overcome this problem, some of *Direction of Arrival* algorithm is used.

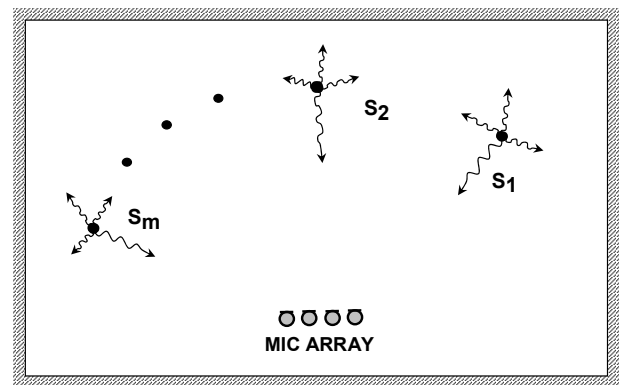


Fig.1. Microphone array in the room with reverberation.

In this paper we proposed different solution which exploits existence of pauses in the natural speech. GSC weightings should be estimated during these pauses in desired speech signal. The weightings estimated by this way could perform maximum interference cancellation with no cancellation of the desired signal nevertheless the existence of reverberation in the room. The reflections of the desired speech source cannot be completely canceled, but residual reflections do not degrade speech intelligibility because human ear is insensitive to low reverberation.

In the section 2 it is shown that ordinary GSC in room with reverberation produces the desired signal cancellation and that the cancellation is proportional to the correlation between the desired signal and its reflections. In the section 3 it is shown that interferences can be completely canceled with no cancellation of the desired signal if GSC weightings are estimated in the pauses of desired signal. In the section 4 experimental tests of the proposed method are discussed. The room with reverberation is simulated using Allen's at al. [6] image method. Using objective cepstral distance measure the following methods are compared: CBF, ordinary GSC, pause

based GSC and GSC with ideally estimated weightings. It was shown that pause based GSC algorithm is superior compared to other methods.

## 2. Full adaptation of GSC in room with reverberation

Suppose that  $m$  acoustic sources and uniform linear microphone array with  $n$  microphones are in the room with reverberation as shown in Fig.1.

Microphone signals can be expressed in DFT domain as

$$\mathbf{H}\mathbf{S} + \mathbf{N} = \mathbf{X}, \quad (1)$$

where  $\mathbf{S}$  is the vector of  $m$  acoustic sources  $\mathbf{S}=[s_1 s_2 \dots s_m]'$  ( $'$  is matrix transpose operator). First signal  $s_1$  is the desired signal, while others are interferences. Transfer function from each source to all microphones is defined by matrix  $\mathbf{H}$

$$\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_m], \quad (2)$$

where  $\mathbf{h}_i$ ,  $i=1, \dots, m$  are vectors of transfer functions from source  $i$  to all microphones.  $\mathbf{N}$  is vector of additive white noise signals and  $\mathbf{X}$  is vector of microphone signals.

Microphone signals are processed with CGS beamformer displayed on Fig.2.

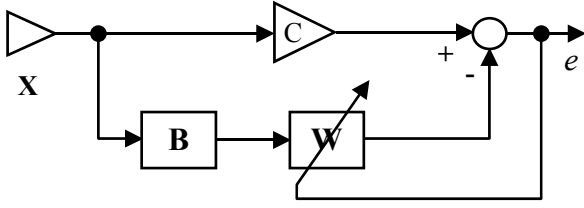


Fig.2. Generalized Sidelobe Canceller.

Output of the GSC signal  $\mathbf{e}$  can be expressed with relation

$$\mathbf{e} = \mathbf{C}^H \mathbf{X} - \mathbf{W}^H \mathbf{B}^H \mathbf{X} \quad (3)$$

where  $\mathbf{C}$  is the steering vector,  $\mathbf{B}$  is blocking matrix and  $\mathbf{W}$  is the vector of GSC weightings. Operator  $^H$  denotes complex conjugate transpose of the matrix/vector. Using minimum variance criterion optimal weightings are estimated with formula [3]

$$\mathbf{W} = (\mathbf{B}^H \Phi_x \mathbf{B})^{-1} \mathbf{B}^H \Phi_x \mathbf{C}, \quad (4)$$

where  $\Phi_x$  is the covariance matrix defined by

$$\Phi_x = E\{\mathbf{X}\mathbf{X}^H\}. \quad (5)$$

The output power is given by

$$\begin{aligned} P_{out} &= E\{\mathbf{e}\mathbf{e}^*\} = (\mathbf{C}^H - \mathbf{W}^H \mathbf{B}^H) \Phi_x (\mathbf{C} - \mathbf{B}\mathbf{W}) = \\ &= \mathbf{C}^H \Phi_x \mathbf{C} - \mathbf{C}^H \Phi_x \mathbf{B} (\mathbf{B}^H \Phi_x \mathbf{B})^{-1} \mathbf{B}^H \Phi_x \mathbf{C} \quad (6) \end{aligned}$$

Transfer function from  $s_1$  to all microphones denoted with  $\mathbf{h}_1$ , can be expressed as a sum of the transfer function of the direct wave  $\mathbf{h}_S$  and sum of transfer functions of all reflections

$\mathbf{h}_R$ . Thus, the contribution of the source  $s_1$  to microphone signals, denoted as  $\mathbf{X}_1$ , can be expressed with relation

$$\mathbf{X}_1 = \mathbf{h}_1 s_1 = (\mathbf{h}_S + \mathbf{h}_R) s_1. \quad (7)$$

The contribution of interferences and noise can be expressed with relation

$$\mathbf{X}_{IN} = \sum_{i=2}^m \mathbf{h}_i s_i + \mathbf{N}. \quad (8)$$

Using relations (7) and (8) and assumption that source  $s_1$  is uncorrelated with interferences and white noise  $\mathbf{N}$ , the covariance matrix  $\Phi_x$  can be decomposed on two components

$$\Phi_x = \Phi_{S+R} + \Phi_{IN}, \quad (9)$$

where

$$\Phi_{S+R} = \Phi_S + \Phi_{S,R} + \Phi_{R,S} + \Phi_R, \quad (10)$$

$$\Phi_S = E\{\mathbf{h}_S s_1 s_1^* \mathbf{h}_S^*\}, \quad \Phi_{S,R} = E\{\mathbf{h}_S s_1 s_1^* \mathbf{h}_R^*\},$$

$$\Phi_{R,S} = E\{\mathbf{h}_R s_1 s_1^* \mathbf{h}_S^*\}, \quad \Phi_R = E\{\mathbf{h}_R s_1 s_1^* \mathbf{h}_R^*\},$$

$$\Phi_{IN} = E\{\mathbf{X}_{IN} \mathbf{X}_{IN}^*\}. \quad (11)$$

\* is conjugate complex operator. Using property of the blocking matrix  $\mathbf{B}^H \mathbf{h}_S = 0$  the following expressions are valid

$$\mathbf{B}^H \Phi_x = \mathbf{B}^H (\Phi_{R,S} + \Phi_{RIN}), \quad (12)$$

$$\Phi_{RIN} = \Phi_R + \Phi_{IN}, \quad (13)$$

$$\mathbf{B}^H \Phi_x \mathbf{B} = \mathbf{B}^H \Phi_{RIN} \mathbf{B}. \quad (14)$$

Using (6), (12), (14), and with some algebraic transformations, the output power can be expressed with relation

$$\begin{aligned} P_{out} &= (\mathbf{C}^H - \tilde{\mathbf{W}}^H \mathbf{B}^H) \Phi_{S+R} (\mathbf{C} - \mathbf{B}\tilde{\mathbf{W}}) + \\ &+ (\mathbf{C}^H - \tilde{\mathbf{W}}^H \mathbf{B}^H) \Phi_{IN} (\mathbf{C} - \mathbf{B}\tilde{\mathbf{W}}) - \\ &- \mathbf{C}^H \Phi_{S,R} \mathbf{B} (\mathbf{B}^H \Phi_{RIN} \mathbf{B})^{-1} \mathbf{B}^H \Phi_{R,S} \mathbf{C} \quad (15) \end{aligned}$$

where

$$\tilde{\mathbf{W}} = (\mathbf{B}^H \Phi_{RIN} \mathbf{B})^{-1} \mathbf{B}^H \Phi_{RIN} \mathbf{C} \quad (16)$$

The first part in (15) is the sum of powers of the desired signal and its reflections. The second part is residual of the interferences and noise. The third part is a squared form and is proportional with the correlation between the desired signal and its reflections. The sign of this part is negative and it represents suppression of the desired signal.

The reflections of the desired signal degrade the desired signal in two ways. The first one is increasing of overall interferences' power. Another degradation is more serious because it suppresses the desired signal. The consequence is a

reduction in the signal to interference ratio. In the following section it will be shown that estimation weightings during pauses of the desired speech signal prevents suppression of the desired signal.

### 3. Parameter estimation during pause of the desired speech signal

In the pause of the desired speech signal only interferences and noise are present. Using relation (4) and taking in to account that  $\Phi_x = \Phi_{IN}$ , weightings can be estimated by formula

$$W_p = (B^H \Phi_{IN} B)^{-1} B^H \Phi_{IN} C \quad (17)$$

After estimation, the weightings are frozen and used to process microphone signals when both desired signal and interferences are present. The power of the output in this case is

$$\begin{aligned} P_{out} &= (C^H - W_p^H B^H) \Phi_{S+R} (C - B W_p) + \\ &+ (C^H - W_p^H B^H) \Phi_{IN} (C - B W_p) = \\ &= \sigma_{S+R}^2 + \sigma_{IN}^2 \end{aligned} \quad (18)$$

The first part of (18) is the desired signal response with its all reflections. Direct wave appears on the output without attenuation, while reflections are partly attenuated. Although reflections are not completely cancelled, they do not reduce speech intelligibility. The reason for this is that human ear is insensitive to small reverberation.

The second part of (18) is the power of residual interferences and noise. This part will tends to zero under following conditions:

1. The white noise level should tend to zero.
2. The number of sources of interferences should be less than the number of microphones.
3. The window of DFT should be larger than room impulse response.

Under these conditions the second part of (18) is nearly zero and the output is a mixture of the desired signal and its attenuated reflections.

### 4. Experimental results

To evaluate theoretical results we simulated a room with reverberation displayed on Fig. 3. Reverberation is simulated with the image method displayed in [6]. The number of sources was 2. Microphone array had 8 microphones. The damping factor of all walls was set to 0.7. The sampling rate of speech signals was 10KHz. The duration of test signals was 10s. To obtain high interference suppression we used DFT with 4096 points.

Quality of the restored signal was measured with LPC cepstral distortion measure, and results are shown in Table 1. In all experiments the output of GSC was compared with two

signals. The first one was  $S_1$  (column 1 of Table 1). The second one was the signal of microphone 1 when only  $S_1$  was active in the room (column 2 of Table 1). The reason why we used the second reference signal is that the proposed method do not provide high cancellation of the reverberation of desired signal. Because of that the restored signal is much more similar to the room response of the  $S_1$  then to the origin  $S_1$ .

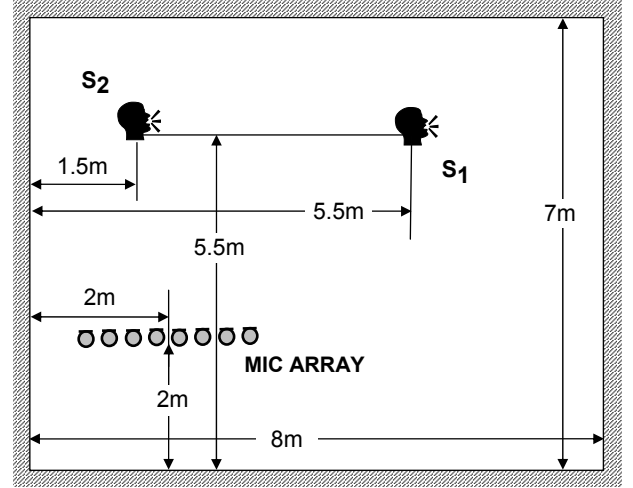


Fig.3. Experimental setup with simulated room with reverberation.

In the first row of the Table 1 there are cepstral distance measures applied to the signal restored by Conventional Beamformer.

The second row of Table 1 shows distance measures for the signal restored with GSC. Its weightings are estimated over the whole signal.

The third row of Table 1 shows distance measures for the signal restored with the proposed method. GSC parameters were estimated during the pauses of desired signal labeled by hand. In the speech signal  $S_1$  there were 4 pauses shorter then 0.5 sec each.

The fourth row of Table 1 shows distance measures for the signal restored with weightings estimated under following (ideal) condition. The addition simulation is performed with source  $S_1$  muted. Then the estimation of GSC weightings was performed. With these weightings signal  $S_1$  was restaured. Obtained speech quality is upper bound of the proposed method.

From Table 1 we can see that the best quality of the restored desired signal is obtained in an ideal case where weightings are estimated without the desired signal. By subjective test the result is even better than the distance measure shows. The reason for this is that human ear is practically insensitive to the residual reverberation. The restored signal  $S_1$  under these conditions is shown in Fig. 4. Good results are also obtained by estimating weightings during pauses of the desired signal. The GSC ordinary applied to the whole signal produces a worse result because of cancellation of the desired signal. The worst result is obtained with the Conventional Beamformer.

Table 1. Cepstral distance measures

Algorithms	Compared with	
	Original $S_1$	$S_1$ in room
CBF	1.182	0.860
(Ordinary) GSC	1.064	0.858
Pause based GSC	0.917	0.621
GSC in ideal case	0.804	0.524

## 5. Conclusions

Unwanted cancellation of the desired signal degrades SNR improvement of the GSC canceller. It is shown that output of the GSC consists of three components. The first component is response to direct wave of the desired signal and all its reflections. The second one is residual of suppressed interferences and white noise. The third one is quadratic form of the correlation between direct wave of desired signal and all its reflections. The sign of this part is negative and it explains cancellation of the desired signal. It was shown that there is no cancellation when estimation of GSC weightings is done during pauses of the desired signal. The experimental tests verified the theoretical results. They also proved significant improvement of the speech quality when GSC weightings were estimated with proposed procedure.

Practical implementation of the proposed method is a two-step algorithm. In the first step pause detection algorithm should be used to label pauses in the desired signal. In the second step optimal parameters have to be estimated on detected pauses to prevent cancellation of the desired signal.

## 6. Acknowledgements

This work was partially supported by the Ministry of Science, Technology and Development under Grant number OI-1784.

## 7. References

- [1] Claude Marro, Yannick Mahieux, and K. Uwe Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering", *IEEE Transactions on Speech and Audio Processing*, 6(3): 240-259, May 1998.
- [2] Frost, O.L. "An algorithm for linearly constrained adaptive array processing", *Proceedings of IEEE*, vol. 60, no.8, pp. 926-935, Aug. 1972.
- [3] L. J. Griffiths, C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming", *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp.27-34, Jan.1982.
- [4] Zoran Šarić, Milorad Cetina, "Partially Adaptive Generalized Sidelobe Canceller with Prescribed Zeros' Positions", *SIGNAL PROCESSING VIII Theories and Applications*, (EUSIPCO-96), pp 947-950, 1996. Trieste Italy.
- [5] Lucas C. Parra, Cristopher V. Alvino, "Geometric Source Separation: Merging Convolutional Source Separation With Geometric Beamforming", *IEEE Transactions on Speech and Audio Processing*, 10(6): 352-362, September 2002.
- [6] Jont B. Allen, David A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Amer.* Vol.65, no.4, pp 943-950, Apr.1979.

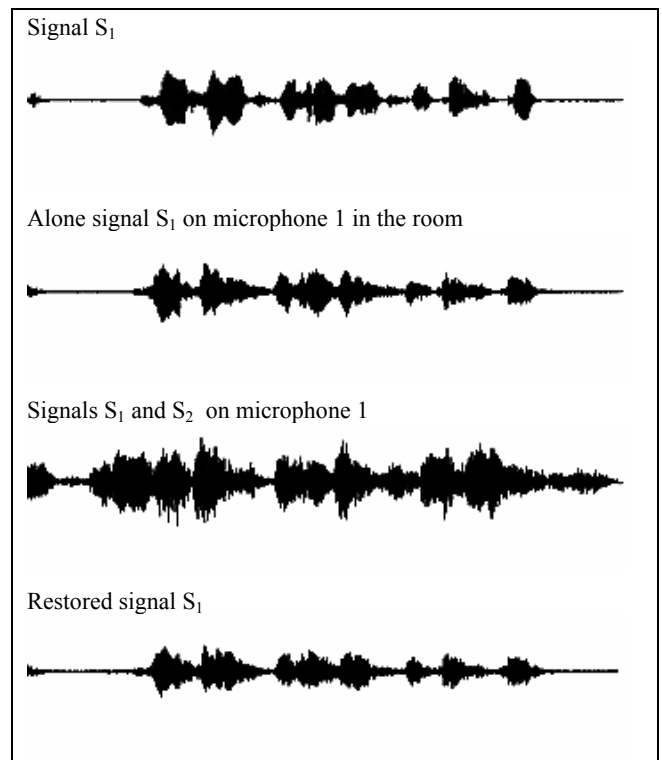


Fig. 4. Restoring signal  $S_1$  in the room with reverberation with GSC weightings estimated on the part when only  $S_2$  is active.