I INTRODUCTION

Microphone arrays are powerful tool for the acoustic interference suppression. They usually utilize adaptive algorithm that adapts its beam pattern so to maximally suppress interferences, having unity gain for the desired signal. Its weightings are usually estimated applying minimum variance (MV) criterion utilizing one of the two beamformer types: Frost’s beamformer [1] or generalized sidelobe canceller (GSC) [2]. The common drawback of MV based adaptive beamformers is theirs sensitivity to the room reverberation, steering and calibration error. Any of these disturbances cause cancellation and distortion of the desired signal.

The great efforts were done to improve the performance of the microphone array and to make it robust against reverberation of the room. Some of the solutions utilize a beamformer type algorithm, followed by postprocessor. Claude Marro et al. [3] suggested splitting the microphone array into differentially equispaced subarrays. The proposed microphone array is “static” (nonadaptive). Its major advantage is robustness against acoustical environment. Unfortunately, its performance in suppression of the coherent noise components is less compared with adaptive algorithms such as GSC.

Zoran Šarić and Milorad Cetina [4] proposed partially adaptive GSC with spatial zeros placed on the directions of the known interference sources. The rest of the unknown interference sources are suppressing by the adaptive algorithm.

Sharon Gannot et al. [5] proposed GSC solution, which is adapted to the general transfer function (TF) case. They suggested estimation of TFs by exploiting the nonstationary characteristics of the desired signal.

In order to reduce desired signal cancellation, Greenberg et al. [6] proposed adaptive algorithm controlled by the cross correlation of the microphone signals. The algorithm is additionally improved by Osamu Hoshuyama et al. [7] utilizing adaptive blocking matrix.

In this paper the influence of the room’s reverberation to the performance of the GSC algorithm will be analyzed using general transfer functions (TF) model [5]. It will be shown that the desired signal cancellation is proportional to the cross correlation of the desired signal with its room reflections [8]. If the GSC weightings are estimated in the pauses of desired signal, there is no cancellation of the desired signal. The appropriate pause detection based algorithm in DFT domain will be proposed and tested by the simulations.

II FULL ADAPTATION GSC

Let us assume that the signals are processing in DFT domain. To simplify notation, we will omit frequency index $f$ in all variables like this $X=X(f)$. Suppose that $m$ acoustic sources and uniform linear microphone array with $n$ microphones are in the room with reverberation. Microphone signals, arranged in vector $X$, are obtained by the following model

$$
H S + N = X,
$$

where $S$ is the vector of $m$ acoustic sources $S=[s_1, s_2, \ldots, s_m]'$. Superscript $'$ denotes matrix transpose operator. Signal $s_i$ is the desired signal, while the others are interferences. $H$ is matrix of transfer functions from each source to all microphones. It is defined by

$$
H = [h_1, h_2, \ldots, h_m],
$$

where $h_i, i=1,\ldots,m$ are vectors of transfer functions from source $i$ to all microphones. TFs from sources to microphones are assumed to be general case [5] that include direct waves and all theirs reflections. $N$ is vector of additive white noise signals.

Microphone signals are processed with GSC beamformer displayed on Fig.1. The output of the GSC can be expressed with relation

$$
e = C^H X - W^H B^H X,
$$

where $C$ is the steering vector, $B$ is blocking matrix and $W$ is the vector of GSC weightings. Superscript $^H$ denotes complex conjugate transpose. We will decompose the vector $h_i$ on two components. The first component $h_{i_1}$ is TF of the direct wave of the desired signal described by the relation
\[ \mathbf{h}_s = \left[ e^{-j2\pi \tau_1}, e^{-j2\pi \tau_2}, \ldots, e^{-j2\pi \tau_n} \right]^H, \]  

(4)

where constants \( \tau_i, i=1, n \) are the direct wave delays to each microphone. The second component \( \mathbf{h}_0 \) is the sum of TFs of the all reflections of the desired signal. Steering vector \( \mathbf{C} \) is defined using direct wave TF 

\[ \mathbf{C} = \mathbf{h}_s \quad . \]  

(5)

Blocking matrix has to satisfy the following condition [5] 

\[ \mathbf{B}^H \mathbf{h}_0 = 0 . \]  

(6)

Using MV criterion GSC weightings \( \mathbf{W} \) can be estimated with formula [1] 

\[ \mathbf{W} = \left( \mathbf{B}^H \mathbf{\Phi}_s \mathbf{B} \right)^{-1} \mathbf{B}^H \mathbf{\Phi}_s \mathbf{C}, \]  

(7)

where \( \mathbf{\Phi}_s \) is covariance matrix defined with 

\[ \mathbf{\Phi}_s = E\{\mathbf{X} \mathbf{X}^H\} \quad . \]  

(8)

The output power \( \mathbf{P}_{out} \) is given by 

\[ \mathbf{P}_{out} = E\{\mathbf{ee}^*\} = C^H \mathbf{\Phi}_s \mathbf{C} - C^H \mathbf{\Phi}_s \mathbf{B} (\mathbf{B}^H \mathbf{\Phi}_s \mathbf{B})^{-1} \mathbf{B}^H \mathbf{\Phi}_s \mathbf{C}. \]  

(9)

The contribution of the desired signal \( s_t \) to the microphone signal vector \( \mathbf{X} \) can be expressed with relation (10) 

\[ \mathbf{X}_t = \mathbf{h}_t s_t = (\mathbf{h}_s + \mathbf{h}_0) s_t. \]  

(10)

The contribution of the interference signals and the noise signal can be expressed with relation (11) 

\[ \mathbf{X}_n = \sum_{i=2}^{\infty} \mathbf{h}_i s_i + \mathbf{N}. \]  

(11)

Using relations (10) and (11) and assumption that the source \( s_t \) is uncorrelated with interferences and the white noise, the covariance matrix \( \mathbf{\Phi}_s \) can be decomposed on two components 

\[ \mathbf{\Phi}_s = \mathbf{\Phi}_{s,R} + \mathbf{\Phi}_{n,R}, \]  

(12)

where 

\[ \mathbf{\Phi}_{s,R} = \mathbf{\Phi}_s + \mathbf{\Phi}_{s,R} + \mathbf{\Phi}_{s,R} + \mathbf{\Phi}_{s,R} , \]

\[ \mathbf{\Phi}_s = E\{\mathbf{h}_s s_s^H \}, \]

\[ \mathbf{\Phi}_{s,R} = \mathbf{\Phi}_{s,N} = E\{\mathbf{h}_s s_s^H \}, \]

\[ \mathbf{\Phi}_{n,R} = E\{\mathbf{h}_n s_s^H \}. \]

(13)

Superscript * denotes conjugate complex operator. Using property of the blocking matrix (6) and taking relations (9), (12), with some algebraic transformations, the output power can be expressed with relation [8] 

\[ \mathbf{P}_{out} = \left( \mathbf{C}^H - \mathbf{W}^H \mathbf{B} \right) \mathbf{\Phi}_{s,R} (\mathbf{C} - \mathbf{B} \mathbf{W}) + \left( \mathbf{C}^H - \mathbf{W}^H \mathbf{B} \right) \mathbf{\Phi}_{n,R} (\mathbf{C} - \mathbf{B} \mathbf{W}) - \mathbf{C}^H \mathbf{\Phi}_{s,R} \mathbf{B} (\mathbf{B}^H \mathbf{\Phi}_{s,R} \mathbf{B})^{-1} \mathbf{B}^H \mathbf{\Phi}_{s,R} \mathbf{C}. \]  

(14)

The first part in (13) is the sum of powers of the desired signal and its reflections. The second part is residual of the interference signals and noise. The third part is a squared form proportional with the cross correlation between the desired signal and its reflections. The sign of this part is negative and represents suppression of the desired signal. In the following section it will be shown that GSC weightings estimated during pauses of the desired speech signal prevents suppression of the desired signal.

### III ADAPTATION DURING PAUSES OF SPEECH

In the pause of the desired speech signal, only interferences and noise are present. Using the relation (7) and taking in to account that \( \mathbf{\Phi}_s = \mathbf{\Phi}_{s,R} \), GSC weightings can be estimated by formula 

\[ \mathbf{W}_p = \left( \mathbf{B}^H \mathbf{\Phi}_{s,R} \right)^{-1} \mathbf{B}^H \mathbf{\Phi}_{s,R} \mathbf{C}. \]  

(15)

After estimation phase, the GSC weightings are frozen and used to process microphone signals. In this case, the output power is 

\[ \mathbf{P}_{out} = \left( \mathbf{C}^H - \mathbf{W}_p^H \mathbf{B} \right) \mathbf{\Phi}_{s,R} (\mathbf{C} - \mathbf{B} \mathbf{W}_p) + \left( \mathbf{C}^H - \mathbf{W}_p^H \mathbf{B} \right) \mathbf{\Phi}_{n,R} (\mathbf{C} - \mathbf{B} \mathbf{W}_p). \]  

(16)

The first part of (16) is the desired signal response with its reflections. Direct wave signal appears on the output with unit gain, while reflections are partly attenuated. The second part of (16) is the power of residual interference and noise. Comparing (16) with (13) we can see that the third part of (13) disappeared. Hence, we can conclude that there is no cancellation of the desired signal.

### III PAUSE DETECTION

To prevent the desired signal cancellation the GSC weightings have to be estimated during pauses of desired signal. Pause detection is based on the estimation of the signal to interference ratio (SIR). The best available approximation of the desired signal \( s_t \) is the GSC output. The best approximation of the interference power is average power of the reference signals \( \mathbf{Z} \). Taking into account the strong correlation of the neighboring frequency bins of the speech signal, the estimation of SIR is obtained by the formula (17).

\[ l(f) = 10 \log \left( \frac{\overline{\mathbf{P}}_s(f)}{\overline{\mathbf{P}}_i(f)} \right), \]  

(17)

where \( \overline{\mathbf{P}}_s(f) \) is frequency smoothed power of the output around frequency bin \( f \) 

\[ \overline{\mathbf{P}}_s(f) = \frac{1}{2\Delta + 1} \sum_{f-\Delta}^{f+\Delta} |\mathbf{Z}(f)|^2. \]

The value \( \overline{\mathbf{P}}_i(f) \) is frequency smoothed average power of the reference signals defined by formula (18) 

\[ \overline{\mathbf{P}}_i(f) = \frac{1}{(n-1)(2\Delta + 1)} \sum_{f-\Delta}^{f+\Delta} |\mathbf{Z}(f)|^2. \]  

(18)

The value \( \Delta \) defines the smoothing window length. Pause detection is performed in every frequency bin independently. The decision rule can be expressed by the following relation 

\[ l(f) \geq \lambda_p(f), \]  

(19)

where \( d_0 \) and \( d_1 \) are decisions of the hypothesis \( H_0 \) (pause) and \( H_1 \) (speech) respectively. \( \lambda_p(f) \) is frequency dependent threshold. The rate of the decision \( d_0 \) depends on the threshold \( \lambda_p(f) \). If the a priori probability of the hypothesis \( H_0 \) is equal to \( q, q=p(H_0), \) it is reasonable to adjust threshold
\lambda_p(f) so that the rate of the decision \(d_0\) is equal \(q\). This can be described by formula
\[
\lambda_p(f) = \lambda \left( p_x(d_0) = q \right),
\]
where \(p_x(d_0)\) is decision \(d_0\) rate under threshold \(\lambda\). The rate of the decision \(d_0\) can be evaluated using histogram \(\text{hist}(l(f))\) of the last \(N\) values of the estimated signal to interference ratio \(l(f)\) by formula
\[
p_x(d_0) = \frac{1}{N} \sum_{t=n}^{t=n-1} \text{hist}(l(t)).
\]
In practice we don’t know the true value of \(q\). Hence, we must assume some value of \(q\) that will meet two requirements: a good pause detection, and small false alarm rate. The adaptation of the GSC weightings is performed by the formula
\[
W^H(f,t) = W^H(f,t-1) - \mu L_p(f,t) Z'(f,t) e^{*}(f,t)
\]
where \(W(f,t)\) is weightings vector at frequency bin \(f\) and time index \(t\), \(L_p(f,t)\) is soft decision between hypothesis \(H_0\) and \(H_1\) by formula
\[
L_p(f,t) = \frac{1}{e^{(\beta (\theta_0(f,t)-\theta_1(f,t)))^2} + 1}
\]
where \(\beta\) is experimentally determined positive constant.

IV EXPERIMENTAL RESULTS

The proposed algorithm has been evaluated by simulation of the room with reverberation using Allen’s image method [9]. The reverberation time was \(T_{60} = 270\) ms (Fig. 2). The number of sources was 2. Microphone array contained 8 microphones spaced 6cm each other. The sampling rate of speech signals was 10 KHz. The duration of the each test signal was 10s. To obtain high interference suppression in the room with long reverberation time, we used DFT with 4096 points. The quality of the desired speech signal restoration was evaluated by cepstral distortion measure, and the results are shown in Table 1.

![Fig.2. 8 microphone array in simulated room with reverberation time 270ms](image.png)

In all experiments the output of GSC was compared with two reference signals. The first one was origin signal \(S_1\) (Fig. 3a), and the second one was the room response of the \(S_1\) (Fig.3b). The reason why we used the second reference signal is that in all of the tested methods the restored signal is much more similar to the room response then to the origin \(S_1\).

In the row 1 of the Table 1, there are cepstral distance measures of the signal restored by conventional beamformer (CBF). In the row 2 there are distance measures of the signal restored with ordinary GSC. Its weightings are estimated over the whole signal. In the row 3 there are distance measures of the signal restored with weightings estimated on hand labeled pauses. In the row 4, there are cepstral distance measures for the signal restored with proposed pause detection based algorithm defined by relations (17)-(23). Finally, GSC weightings were estimated under idealistic case when signal \(S_1\) was muted and only interference signal \(S_2\) was present. The distance measures for this case are in the row 5.

<table>
<thead>
<tr>
<th>Estimation method</th>
<th>Compared with</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>origin (S_1)</td>
</tr>
<tr>
<td>1.</td>
<td>CBF</td>
</tr>
<tr>
<td>2.</td>
<td>Full adaptation GSC</td>
</tr>
<tr>
<td>3.</td>
<td>Pauses labeled by hand</td>
</tr>
<tr>
<td>4.</td>
<td>Proposed algorithm</td>
</tr>
<tr>
<td>5.</td>
<td>Estimation in absence of (S_1)</td>
</tr>
</tbody>
</table>

We can shortly summarize results from the Table 1. As expected, the worst result is obtained with the conventional beamformer. The better result is obtained with ordinary GSC, but the restored signal is evidently degraded due to signal cancellation. Further signal improvement is obtained with GSC weightings estimated on the hand labeled pauses. Some more additional improvement is obtained used proposed pause detection algorithm. The reason for this improvement is that the proposed pause detector takes pauses over separate spectral components. The best quality of the restored signal is obtained in an ideal case where weightings are estimated in absence of the desired signal. It is actually the upper limit of the restoring performance.

V CONCLUSIONS

It is analytically shown that reverberation of the room cause cancellation of the desired signal. The level of the desired signal cancellation is proportional to the cross correlation of the direct wave and reflections of the desired signal. It is also shown that there is no cancellation of the desired signal when estimation of GSC weightings is performed during pauses of the desired signal.

The appropriate pause detection algorithm is proposed to prevent adaptation during speech activity. The superiority of the proposed algorithm compared to CBF and ordinary GSC algorithm is experimentally proved.

REFERENCES


Acknowledgements: This work was partly supported by Ministry of Science, Technology and Development of the Republic Serbia under Grant number OI-1784.

Abstract: Generalized sidelobe canceller (GSC), that exploits minimum variance (MV) criterion, is inefficient in interference suppression when there is any correlation between interference and desired signal. The reason for this is unwanted cancellation and distortion of desired signal. In this paper the performance of GSC in room with reverberation is analyzed. It was shown that cancellation of the desired signal is proportional to the correlation between the direct wave of desired signal and its reflections. It is also shown that there is no cancellation if the weightings of the GSC are estimated in pauses of desired speech signal. Using these facts, the algorithm based on pause detection was proposed. Simulations of the room with reverberation prove the advantage of proposed algorithm.

ADAPTIVNI MIKROFONSKI NIZ U PROSTORJI SA REVERBERACIJOM,
Zoran M. Šarić, Slobodan T. Jovičić

Fig. 3. Signal restoring. a) origin signal S1, b) signal S1 recorded in room with reverberation, c) mixture of signals S1 and S2 on microphone 1, d) S1 restored by ordinary GSC, e) S1 restored by GSC weighting estimated on segments without S1.