OCTOBER 01 2005

Application of the maximum signal to interference ratio criterion to the adaptive microphone array

Slobodan T. Jovicic; Zoran M. Saric; Srbijanka R. Turajlic

ARLO 6, 232–237 (2005)
https://doi.org/10.1121/1.1989785

Application of the maximum signal to interference ratio criterion to the adaptive microphone array

Slobodan T. Jovicic
School of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia and Montenegro
jovicic@etf.bg.ac.yu

Zoran M. Saric
Institute of Security, Kraljice Ane bb, Belgrade, Serbia and Montenegro
sare@yubc.net

Srbijanka R. Turajlic
School of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia and Montenegro
sturajlic@aaen.edu.yu

Abstract: The minimum variance (MV) criterion is widely used for the weight vector estimation of the adaptive microphone array. The drawback of this criterion is the cancellation of the desired speech signal and its degradation in multipath wave propagation environment. Applying the maximum signal to interference ratio (MSIR) instead of the MV criterion has two benefits. The first one is the high suppression of the interferences and the second one the desired speech enhancement. The proposed MSIR criterion is applied to the new generalized eigenvalue based beamformer (GEVBF). Its superiority is experimentally proved by simulating a room with reverberation.

PACS numbers: 43.60.Fg, 43.60.Mn, 43.72.Ew
Date Received: March 10, 2005 **Date Accepted:** June 30, 2005

1. Introduction

The problem of high quality speech recording in a room with reverberation and the cocktail-party interference has long been under consideration. It has been established that microphone arrays, compared to a single microphone, render a better quality of speech recording. The reason being that is that they usually adapt their beam pattern so as to maximally suppress interferences, while maintaining the unity gain for the desired speech signal. The minimum variance (MV) criterion is commonly used to estimate the weight vector of microphone array.[1,2] The common drawback of the MV-based adaptive microphone arrays is their sensitivity to the room reverberation that causes the cancellation of the desired speech signal.[3–5] This cancellation can be prevented if the weight vector is estimated within pauses of the desired speech.[5,6]

The results presented in this paper will demonstrate that the quality of the restored speech signal can be further improved by applying the maximum signal to interference ratio (MSIR) criterion for weight vector estimation. This criterion maximizes the desired signal power, while minimizing the interference power. Applying the MSIR criterion calls for the estimation of the two covariance matrices.[7] The first one, denoted as the *interference matrix*, has to be estimated within the pauses of the desired speech signal. The second one, denoted as the *signal matrix*, has to be estimated when the signal to interference ratio is high. The weight vector is estimated by solving the generalized eigenvalue problem for the signal and interference matrix pair.[7] The proposed MSIR criterion was experimentally verified by the new generalized eigenvector-based beamformer (GEVBF) in a simulated room with reverberation. The comparison to other MV-based algorithms proved the superiority of the proposed algorithm.

 1529-7853/05/6(4)/232/6/$22.50

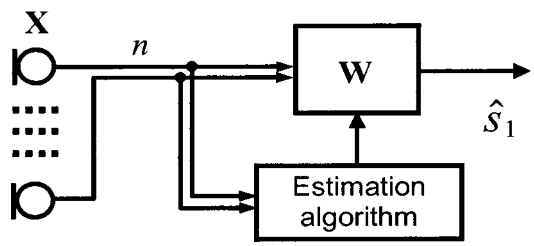

Fig. 1. General structure of the adaptive beamformer.

2. Weight vector estimation

The microphone signals are processed in the DFT domain. All signals are represented by complex DFT coefficients with central frequency f. For the sake of simplicity, the index f will be omitted, i.e., $x(f)=x$. Every DFT bin is processed independently. Let us assume there is an array with n microphones in a room with reverberation. Column vector $\mathbf{X}$ of the n microphone signals can be expressed by

$$\mathbf{X}=\mathbf{S}+\mathbf{U}, \tag{1}$$

where the vector $\mathbf{S}$, $\mathbf{S}=\mathbf{h}_1 s_1$, is the room response to the desired signal s_1, with $\mathbf{h}_1$ being the vector of transfer functions from the desired source s_1 to each of the microphones. The vector of the transfer functions $\mathbf{h}_1$ describes both the direct wave path and the reflections from the walls. Vector $\mathbf{U}$ is the room response to all acoustical interferences.

Microphone signals are processed by an adaptive algorithm displayed in Fig. 1. The estimation of the desired signal $\hat{s}_1$ is the weighed sum of the microphone signals, expressed as

$$\hat{s}_1=\mathbf{W}^*\mathbf{X}, \tag{2}$$

where $\mathbf{W}$ is the weight vector, while superscript * denotes a complex conjugate transpose. The power of the desired signal and the interference power at the beamformer's output are expressed, respectively, as

$$P_{\mathrm{s}}=\mathbf{W}^*\mathbf{\Phi}_{\mathrm{s}}\mathbf{W}, \tag{3}$$

and

$$P_u=\mathbf{W}^*\mathbf{\Phi}_u\mathbf{W}, \tag{4}$$

where $\mathbf{\Phi}_{\mathrm{s}}=E\{\mathbf{S}^*\mathbf{S}\}$ is the *signal matrix* and $\mathbf{\Phi}_u=E\{\mathbf{U}^*\mathbf{U}\}$ is the *interference matrix*. The adaptive algorithm should adjust $\mathbf{W}$ so as to maximize the desired signal power at the output, while minimizing the output interference power. A possible criterion function, which encompasses these two goals, can be formulated as the signal to interference power ratio (SIR),

$$\xi_0=\frac{P_s}{P_u}=\frac{\mathbf{W}^*\mathbf{\Phi}_s\mathbf{W}}{\mathbf{W}^*\mathbf{\Phi}_u\mathbf{W}}. \tag{5}$$

However, estimating $\mathbf{W}$ by maximizing ξ_0 might not be satisfactory because it is highly sensitive on the estimation errors of $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_u$. The estimation error of the $\mathbf{\Phi}_u$ might yield the poorly conditioned solution. This problem can be avoided by adding the term $\beta\mathbf{I}$ to the estimated matrix $\mathbf{\Phi}_u$, $\bar{\mathbf{\Phi}}_u=\beta\mathbf{I}+\mathbf{\Phi}_u$, where $\mathbf{I}$ is the unit matrix and β, $\beta>0$, is a scalar by which a compromise between stability and high interference suppression can be achieved.[7]

On the other hand, the sensitivity of $\mathbf{W}$ to the $\mathbf{\Phi}_s$ estimation error can be reduced by adding the term $\alpha\mathbf{h}_d^*\mathbf{h}_d$ to the estimated signal matrix by

1529-7853/05/6(4)/233/6/$22.50

$$\bar{\Phi}_s = \mathbf{\Phi}_s + \alpha \mathbf{h}_d^* \mathbf{h}_d, \tag{6}$$

with $\mathbf{h}_d$ being the known direct wave transfer function defined by

$$\mathbf{h}_d = [1 \;\; e^{-j2\pi f\tau} \;\cdots\; e^{-j2\pi f(n-1)\tau}]^*, \quad \tau = \frac{d\sin(\theta)}{c}, \tag{7}$$

where θ is the arriving angle of the desired signal, f is the central frequency of the DFT bin, c is the sound velocity and d distance between adjacent microphones. The optimal value of scalar α, $\alpha > 0$ tends to zero if the estimation error of $\mathbf{\Phi}_s$ becomes negligible. In this case, the error of the desired source position is insignificant (6).

Summing up all the arguments, the new SIR criterion ξ_1 becomes

$$\xi_1 = \frac{\mathbf{W}^*(\mathbf{\Phi}_s + \alpha \mathbf{h}_d \mathbf{h}_d^*)\mathbf{W}}{\mathbf{W}^*(\mathbf{\Phi}_u + \beta \mathbf{I})\mathbf{W}}. \tag{8}$$

The maximum of the ξ_1 (MSIR) is achieved with an eigenvector corresponding to the largest eigenvalue of the generalized eigenvalue problem,[7]

$$(\mathbf{\Phi}_s + \alpha \mathbf{h}_d \mathbf{h}_d^*)\mathbf{W} = \lambda(\mathbf{\Phi}_u + \beta \mathbf{I})\mathbf{W}. \tag{9}$$

Weight vector $\mathbf{W}$ estimated by (9) defines generalized eigenvalue beamformer (GEVBF). Matrices $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_u$ have to be estimated from the available data sequence $\mathbf{X}$. The interference matrix $\mathbf{\Phi}_u$, can be estimated within pauses of the desired speech that can be detected by the pause detector.[4–6] Since interference is almost always present, an estimation of the matrix $\mathbf{\Phi}_s$ is somewhat more complicated. A reasonable approach might be to select time intervals where the signal-to-interference ratio is high and use these intervals for estimating the $\mathbf{\Phi}_s$. Finally, the choice of the scalar α also depends on the signal-to-interference ratio. If the signal-to-interference ratio is small, there is a significant error in the estimated matrix $\mathbf{\Phi}_s$, and the scalar α has to be increased.

3. Phase shift compensation

The criterion function ξ_1 is not sensitive to the phase shift in the vector $\mathbf{W}$. Namely, if the $\mathbf{W}_m$ is the solution of the maximization of the ξ_1, then a phase shifted vector $\tilde{\mathbf{W}}_m$, $\tilde{\mathbf{W}}_m = \exp(-j\varphi)\mathbf{W}_m$ is also a solution. The shift angle φ is a random variable in every frequency bin. In order to prevent speech degradation caused by random φ, the phase correction has to be applied so as to obtain the zero phase shift angle for the direct wave. The phase corrected weight vector $\bar{\mathbf{W}}$ can be defined as

$$\bar{\mathbf{W}} = \frac{\mathbf{W}^*\mathbf{C}}{|\mathbf{W}^*\mathbf{C}|}\mathbf{W}, \quad \mathbf{C} \equiv \mathbf{h}_d. \tag{10}$$

4. Experimental results

The proposed GEVBF algorithm with MSIR criterion has been verified in a room with reverberation simulated by Allen's image method.[8] The reverberation time was $T_{60} = 270$ ms. The number of sources was 2: source s_1 was the desired speaker and source s_2 was the interference (Fig. 2). In experiment 1 the interference s_2 was at position (s_2') while in experiment 2 it was at position (s_2''). The microphone array consisted of 8 microphones 6 cm apart. The sampling rate of the speech signals was 10 kHz. The duration of each test signal was 10 s. To obtain high interference suppression in a room with long reverberation time, a DFT with 4096 points was used.

The following algorithms were compared:

 1529-7853/05/6(4)/234/6/$22.50

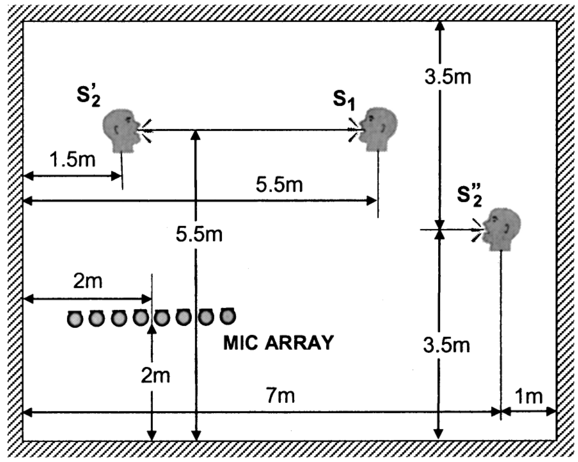

Fig. 2. Experimental setup: simulated room with a reverberation time of 270 ms and an 8 microphone array.

(1) The conventional beamformer (CBF)
(2) Ordinary generalized sidelobe canceller (GSC) with full adaptation[2]
(3) GSC with weights estimated within hand-labeled pause intervals[3]
(4) GSC with weights estimated under an ideal scenario where only interference is present[3]
(5) The proposed GEVBF algorithm with covariance matrices $\mathbf{\Phi}_{\mathrm{s}}$ and $\mathbf{\Phi}_{\mathrm{u}}$ estimated within hand-labeled time intervals of speech and pause, respectively. Parameter α is set to zero as in Ref. 7
(6) The same as (5), except that α=0.005
(7) The proposed GEVBF algorithm with covariance matrices $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_u$ estimated under an ideal scenario, where either a speech signal or interference is exclusively present, with α=0.005
(8) The same as (7), except that α=0.

In algorithms (4), (5), (6), (7), and (8), there was β=0.01. The signal s_1 restored by different algorithms in experiment 1 is depicted in Fig. 3. The quality of the speech signal restoration was evaluated by the cepstral distortion measure, and the results are presented in Table 1. As was expected, the worst result is obtained with a CBF algorithm. A better result is obtained by the full adaptation GSC, but the restored signal is obviously degraded due to signal cancellation. Further improvement is obtained by GSC weights estimated within the hand-labeled pauses (Table 1, row 3). It should be pointed out that the best achievable quality by the MV criterion is under the ideal scenario where the desired signal is muted and only interference is present (Table 1, row 4).

Yet even this result is surpassed by applying the proposed GEVBF algorithm with MSIR criterion, with covariance matrices $\mathbf{\Phi}_{\mathrm{s}}$ and $\mathbf{\Phi}_{\mathrm{u}}$ estimated within hand-labeled time intervals of high speech power and pause in speech, respectively. By applying α=0.005 (Table 1, row 6) we obtain the better result compared to Morgan's method with α=0 (Table 1, row 5). The reason for this is that additional term $\alpha\mathbf{h}_d^*\mathbf{h}_d$ makes the estimation algorithm more robust against an estimation error of $\mathbf{\Phi}_{\mathrm{s}}$.

Naturally, the best quality is obtained by applying the proposed GEVBF algorithm with covariance matrices $\mathbf{\Phi}_s$ and $\mathbf{\Phi}_u$ estimated under an ideal scenario, where either a speech signal or interference is exclusively present (Table 1, rows 7 and 8). As the estimation error of $\mathbf{\Phi}_s$ is negligible, the best result is obtained with α=0 (Morgan's method[7]).

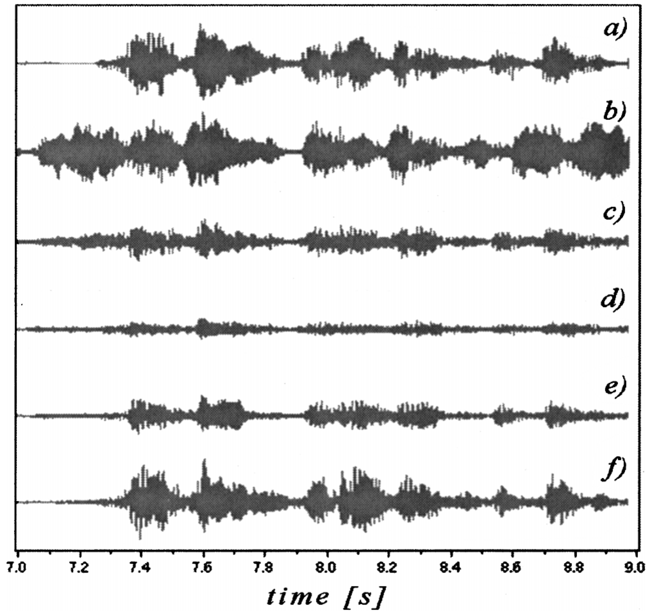

Fig. 3. Time diagrams for the experiment 1: (a) Desired speech signal s_1, (b) signal recorded on microphone 1, (c), (d), (e), and (f) the signal s_1 restored by algorithms (1), (2), (3), and (6), respectively, defined in Sec. 4.

5. Conclusions

In this paper, a new GEVBF algorithm for cocktail party interference suppression based on the maximum signal to interference ratio (MSIR) criterion is proposed. The beamformer weights are estimated by solving a generalized eigenvalue problem involving signal and interference covariance matrices.[7]

The quality of the signal restored by the proposed GEVBF algorithm is better compared to algorithms based on the MV criterion. This improvement is due to the additional knowledge of the signal covariance matrix. This matrix is estimated from microphone signals within time intervals with a high signal-to-interference ratio. The pause detection and the detection of time intervals with a high signal-to-interference ratio can be performed by the algorithm described in Refs. 5 and 6.

Table. 1. Cepstral distortion measures of restored signal.

Estimation algorithms	Cepstral distortion measure	
	Experiment 1	Experiment 2
1. CBF	0.860	1.134
2. GSC	0.758	0.984
3. GSC hand-labeled pauses	0.607	0.800
4. GSC ideal scenario	0.524	0.638
5. GEVBF—hand-labeled intervals (α=0)[a]	0.605	0.612
6. GEVBF—hand-labeled intervals (α=0.005)	0.479	0.545
7. GEVBF—ideal scenario (α=0.005)	0.453	0.506
8. GEVBF—ideal scenario (α=0)[a]	0.419	0.469

[a]Condition like in Morgan's method (Ref. 7).

 1529-7853/05/6(4)/236/6/$22.50

Acknowledgments

This work was partially supported by the Ministry of Science and Environment Protection of the Republic Serbia under Grant No. OI-1784. The authors would like to thank the two unknown reviewers for their helpful and useful comments that significantly improved the quality of the paper.

References and links

[1] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**, 926–935 (1972).
[2] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. **AP-30**, 27–34 (1982).
[3] Z. Saric and S. Jovicic, "Adaptive beamforming in room with reverberation," *Eurospeech 2003*, Geneva, Sep. 2003, pp. 529–532.
[4] O. Hoshuyama, B. Begasse, A. Sugiyama, A. Hirano, "A realtime robust adaptive microphone array controlled by an SNR estimate," *Proc. ICASSP98*, 1988, pp. 3605–3608.
[5] Z. M. Saric and S. T. Jovicic, "Adaptive microphone array based on pause detection," ARLO **5**, 68–74 (2004).
[6] Z. Saric and S. Jovicic, "Subband pause in speech signal detection using microphone array in room with reverberation," *Proceedings of the SPECOM 2004*, St. Petersburg, October 2004, pp. 132–137.
[7] D. R. Morgan, "Adaptive algorithms for solving generalized eigenvalue signal enhancement problem," Signal Processing **84**, 957–968 (2004), ISSN 0165-1684.
[8] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).

