# Adaptive microphone array for unknown desired speaker's transfer function

Istvan I. Papp, Zoran M. Saric, Slobodan T. Jovicic, et al.

Citation: The Journal of the Acoustical Society of America **122**, EL44 (2007); doi: 10.1121/1.2749077
View online: https://doi.org/10.1121/1.2749077
View Table of Contents: https://asa.scitation.org/toc/jas/122/2
Published by the Acoustical Society of America

# Adaptive microphone array for unknown desired speaker's transfer function

**Istvan I. Papp**
*Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia*
*istvan.papp@micronasnit.com*

**Zoran M. Saric**
*MicronasNIT, Fruškogorska 11a, 21000 Novi Sad, Serbia*
*zoran.saric@micronasnit.com*

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

**Nikola Dj. Teslic**
*Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia*
*nikola.teslic@micronasnit.com*

**Abstract:** The main drawback of minimum variance distortionless response (MVDR) beamformer is the cancellation of the desired speech signal and its degradation in multi-path wave propagation environment. To make the adaptive algorithm robust against room reverberation and to prevent desired signal cancellation an estimation of unknown desired speaker's transfer function was proposed. The estimation is based on the signal and the interference covariance matrices. The estimated transfer function is then applied to the MVDR beamformer. The proposed algorithm was tested on a simulated room with reverberation. The results showed better quality of the restored speech compared to some typical adaptive algorithms.



**PACS numbers:** 43.60.Fg, 43.60.Mn, 43.72.Dv [JC]
**Date Received:** April 7, 2007 **Date Accepted:** April 30, 2007

## 1. Introduction

The problem of high quality speech recording in a room with reverberation and the cocktail-party interference has been long under consideration. It has been established that microphone arrays, compared to a single microphone, render a better quality of speech recording. The commonly used minimum variance distortionless response (MVDR) beamformer is the optimal estimator for the Gaussian process and for the known desired signal transfer function.[1–3] In the reverberant room, actual transfer function is not known. The incomplete knowledge of the transfer function causes desired speaker cancellation.[4,6] In order to reduce this cancellation, some linear and quadratic constraints have to be applied.[3] Unfortunately, these constraints can degrade the interference suppression performance significantly. The alternative methods exploit the nonstationary nature of the speech signal[4–7] by estimating array weights during the pauses in the speech. In this case there is no desired speech cancellation.[6]

In this paper a two step minimum variance beamforming algorithm, denoted by TS-MV, is proposed. In the first step the unknown transfer function of the desired speaker is estimated using both estimates of the signal and interference covariance matrices. In the second step the estimated transfer function is used for MVDR weights calculation to prevent desired signal cancellation. In addition, it is shown that the proposed estimation of the transfer function is robust against imperfect signal covariance matrix. The proposed estimation algorithm is experimentally tested in a simulated room with reverberation. Experimental results showed the improvement in restored speech quality compared to some similar algorithms.

 

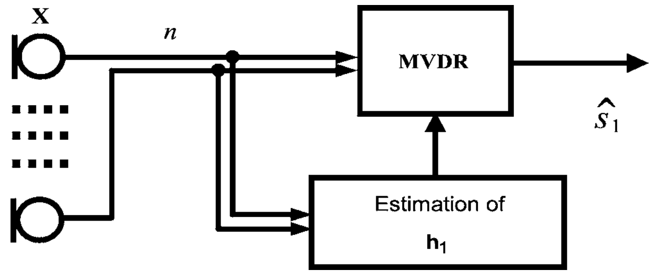


Fig. 1. General structure of the adaptive beamformer.

## 2. Baseline approach

Let us assume a reverberant room with an array of $n$ microphones, desired signal $s_1$, and $m$ acoustical interferences $s_2, \ldots, s_{m+1}$. The microphone signals are processed in discrete Fourier transform (DFT) domain. All signals are represented by the complex DFT coefficients with central frequency $f$. For the sake of simplicity the index $f$ will be omitted, i.e., $x=x(f)$. Column vector $\mathbf{X}$ of the $n$ microphone signals can be expressed by

$$\mathbf{X} = \mathbf{S} + \mathbf{U}, \quad \mathbf{S} = \mathbf{h}_1 s_1, \tag{1}$$

where $n$-column vector $\mathbf{S}$ is the room response to the desired signal $s_1$ excitation, and $n$-column vector $\mathbf{h}_1$ is its transfer function containing both direct path and reflections. The vector $\mathbf{U}$ is the sum of responses to the interference signal vector $\mathbf{S}_I$, $\mathbf{S}_I=[s_2, \ldots, s_{m+1}]'$ and to the uncorrelated microphone noise $\mathbf{N}$, $\mathbf{N}=[n_1 \ldots n_n]'$ expressed by

$$\mathbf{U} = \mathbf{H}_I \mathbf{S}_I + \mathbf{N}, \tag{2}$$

where $\mathbf{H}_I$ is $n \times m$ interference transfer matrix. In the rest of the paper superscript $^H$ denotes a complex conjugate transpose, $^*$ denotes complex conjugation, $'$ denotes matrix/vector transposition, and $E\{\cdot\}$ denotes the statistical expectation operator. Microphone signals are processed by adaptive algorithm displayed in Fig. 1. Output signal $\hat{s}_1$ is the weighted sum of the microphone signals, $\hat{s}_1 = \mathbf{W}^H \mathbf{X}$, where $\mathbf{W}$ is weight vector of the MVDR beamformer expressed by[1,2]

$$\mathbf{W} = \frac{\mathbf{\Phi}_{U,U}^{-1} \mathbf{h}_1}{\mathbf{h}_1^H \mathbf{\Phi}_{U,U}^{-1} \mathbf{h}_1}. \tag{3}$$

The interference cross-spectral matrix $\mathbf{\Phi}_{U,U}$, $\mathbf{\Phi}_{U,U}=E\{\mathbf{U}\mathbf{U}^H\}$ has to be estimated from available measurements $\mathbf{X}$ during the absence of desired speech.[6] The problem is that transfer vector $\mathbf{h}_1$ is not known in reverberant environment. The use of the direct path transfer vector instead of $\mathbf{h}_1$ causes unwanted desired speech cancellation.[4,6] To prevent this, the actual $\mathbf{h}_1$ has to be estimated.

## 3. Transfer function estimation

Transfer vector $\mathbf{h}_1$ can be estimated from the signal covariance matrix $\mathbf{\Phi}_{S,S}$ that is defined by $\mathbf{\Phi}_{S,S}=E\{\mathbf{X}\mathbf{X}^H\}$ under the assumption that only desired signal $s_1$ and noise $\mathbf{N}$ are present. From Eqs. (1) and (2), it follows

$$\mathbf{\Phi}_{S,S} = \Phi_{s,s} \mathbf{h}_1 \mathbf{h}_1^H + \Phi_{N,N} \mathbf{I}, \tag{4}$$

where $\Phi_{s,s}$ is desired signal power, and $\Phi_{N,N}$ is uncorrelated noise power. Using the principal eigenvector $\mathbf{v}_p$ of $\mathbf{\Phi}_{S,S}$, the estimate of $\mathbf{h}_1$ is

$$\hat{\mathbf{h}}_1 = C_\varphi \mathbf{v}_p, \quad C_\varphi = \exp(-j\varphi), \tag{5}$$

where $C_\varphi$ is a unit magnitude complex multiplier that influences only the signal delay.[7] The phase compensation will be defined in Sec. 4. There is a problem in signal covariance matrix estimation because at least one of the interferences is almost always present.[8] Hence, we must take into account that $\mathbf{\Phi}_{S,S}$ is contaminated with $\mathbf{\Phi}_{U,U}$ by

$$\hat{\mathbf{\Phi}}_{S,S} = \alpha \mathbf{\Phi}_{S,S} + (1-\alpha)\mathbf{\Phi}_{U,U} \quad 0.5 < \alpha < 1, \tag{6}$$

where $\alpha$ is a positive scalar. The second term in Eq. (6) significantly degrades the estimation of $\mathbf{h}_1$. The improved estimate of $\mathbf{h}_1$ can be defined under the following assumptions:

(A1)
The estimate of the interference covariance matrix $\mathbf{\Phi}_{U,U}$ is available.
(A2)
The number of the interference signals is less than the number of microphones $(m<n)$.
(A3)
Uncorrelated noise power $\Phi_{N,N}$ is much less than the desired signal power $\Phi_{N,N} \ll \Phi_{s,s}$.

Let us define auxiliary matrix $\mathbf{\Phi}$, $\mathbf{\Phi} = \hat{\mathbf{\Phi}}_{S,S}\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\hat{\mathbf{\Phi}}_{S,S}$. Under assumptions A1, A2, A3, the principal eigenvector of $\mathbf{\Phi}$ can be used as an approximation of the principal eigenvector of $\mathbf{\Phi}_{S,S}$, required in Eq. (5). The proof is given in Appendix A 1.

## 4. Proposed algorithm

Finally, the proposed two step minimum variance (TS-MV) algorithm can be described by:

**Step 1**: Estimate of $\mathbf{h}_1$

(i) Estimate $\mathbf{\Phi}_{U,U}$ on pause intervals of desired speech signal, and $\mathbf{\Phi}_{S,S}$ on intervals with high speech to interference ratio. Calculate $\hat{\mathbf{\Phi}}_{S,S}$ by Eq. (6).

(ii) Calculate auxiliary matrix $\mathbf{\Phi}, \mathbf{\Phi} = \hat{\mathbf{\Phi}}_{S,S}\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\hat{\mathbf{\Phi}}_{S,S}$, calculate principal eigenvector $\mathbf{v}_p$ of the matrix $\mathbf{\Phi}$, and estimate transfer vector $\mathbf{h}_1$ by Eq. (5).

**Step 2**: Apply MVDR

(iii) Apply diagonal loading to the interference covariance matrix $\mathbf{\Phi}_{U,U}$ by $\widetilde{\mathbf{\Phi}}_{U,U} = (\mathbf{\Phi}_{U,U} + \beta \mathbf{I})$, to make the MVDR beamformer robust against steering error and room reverberation.[3,6] Scalar $\beta, \beta > 0$, makes a compromise between stability and high interference suppression.

(iv) Calculate MVDR weight vector $\mathbf{W}$ as $\mathbf{W} = \widetilde{\mathbf{\Phi}}_{U,U}^{-1}\hat{\mathbf{h}}_1 \big/ \hat{\mathbf{h}}_1^H \widetilde{\mathbf{\Phi}}_{U,U}^{-1}\hat{\mathbf{h}}_1$.

(v) As the estimated $\mathbf{h}_1$ has random phase shift factor $C_\varphi$ (5), apply phase compensation by[7] $\bar{\mathbf{W}} = \left(\mathbf{W}^H \mathbf{h}_d / \mathbf{W}^H \mathbf{h}_d\right)\mathbf{W}, \mathbf{h}_d = [1 \; e^{-j2\pi f\tau} \dots e^{-j2\pi f(n-1)\tau}]$, $\tau$ delay on adjacent microphones, where $\mathbf{h}_d$ is the direct path transfer vector.

## 5. Experimental results

The proposed TS-MV algorithm has been examined in a room with reverberation simulated by Allen's image method.[8] The room 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 the experiment 1 the interference $s_2$ was at position $s_2'$ (easier to suppress) while in the experiment 2 it was at position $s_2''$ (harder to suppress). Critical distance boundary was calculated from the room model for which the direct path power is equal to the reverberant power. The microphone array consisted of eight microphones with equidistant spacing of 6 cm. The sampling rate of the speech signals was 10 kHz, while the length of data processing block was 2048 points. Signal of the microphone 1 for position $s_2'$ is in Mm. 1. The following algorithms were compared: (1) The conventional beamformer (CBF) (Mm. 2), (2) Generalized sidelobe canceller (GSC) (Mm. 3),

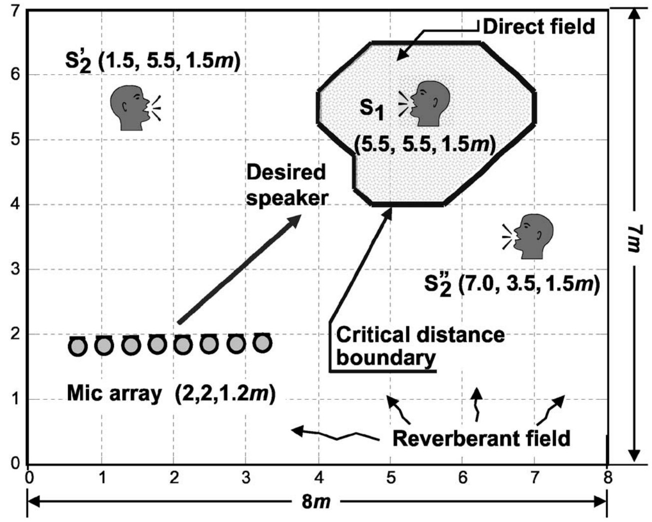


Fig. 2. Simulated room with a reverberation time of 270 ms and a microphone array with eight microphones.

(3) GSC with weights estimated within hand labeled pause intervals,[6] (4) GSC with weights estimated under an ideal scenario where only interference is present,[6] (5) GEVBF with hand-labeled signal and pause intervals,[7] (6) GEVBF with signal and interference covariance matrices estimated under an ideal scenario where either a desired speech or interference is exclusively present, (7) Proposed TS-MV algorithm with covariance matrices $\mathbf{\Phi}_{S,S}$ and $\mathbf{\Phi}_{U,U}$ estimated within hand labeled time intervals of speech and pause, respectively (Mm. 4), (8) Proposed TS-MV algorithm with covariance matrices $\mathbf{\Phi}_{S,S}$ and $\mathbf{\Phi}_{U,U}$ estimated under an ideal scenario, where either a speech signal or interference is exclusively present.

In all algorithms, except CBF, the diagonal loading of the interference covariance matrix is applied to reduce desired signal cancellation.[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 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 MVDR criterion is under the ideal scenario where the desired signal is muted and only interference is present (Table 1, row 4). The additional improvement is obtained by the GEVBF algorithm that maximizes signal to noise ratio.[7] The best results are obtained by the proposed TS-MV algorithm.

Table 1. Cepstral distortion measures of restored signal.

| | Cepstral distortion measure | |
|---|---|---|
| Estimation algorithms | Experiment 1 | Experiment 2 |
| 1. CBF | 0.860 | 1.134 |
| 2. Ordinary GSC with diagonal loading | 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 | 0.479 | 0.545 |
| 6. GEVBF – ideal scenario | 0.453 | 0.506 |
| 7. TS-MV – hand-labeled intervals | 0.414 | 0.427 |
| 8. TS-MV – ideal scenario | 0.362 | 0.369 |

Mm. 1 Microphone 1 signal for position $s_2'$ on Fig. 2 (201 kb). This is a file of type "wav."

Mm. 2 CBF output (209 kb). This is a file of type "wav."

Mm. 3 Output of the MVDR (GSC) with diagonal loading (209 kb). This is a file of type "wav."

Mm. 4 Output of the proposed TS-MV algorithm (209 kb). This is a file of type "wav."

## 6. Conclusions

In this paper, a two step minimum variance (TS-MV) algorithm for acoustical interference suppression in reverberant environment is proposed. In the first step the unknown desired speaker's transfer function is estimated while this estimate is then used for MVDR beamformer in the second step. This estimate reduces cancellation of the desired speaker signal, while at the same time preserves high noise suppression.

An improved estimate of the unknown transfer function is obtained using both signal and interference covariance matrices. The proposed estimation algorithm is robust against imperfect signal covariance matrix estimation. An additional robustness of the algorithm is obtained by diagonal loading of the interference covariance matrix. Tests in the simulations of the room with reverberation proved the superior performance of the algorithm.

## Acknowledgments

This work was partially supported by Ministry of Science and Environment Protection of the Republic Serbia under Grant No. OI-1784.

## Appendix A1 Approximation of the principal eigenvector of $\hat{\mathbf{\Phi}}_{S,S}$

Taking into account Eq. (6) the auxiliary matrix $\mathbf{\Phi}$ can be expressed by

$$\mathbf{\Phi} = \hat{\mathbf{\Phi}}_{S,S}\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\hat{\mathbf{\Phi}}_{S,S} = [\alpha\mathbf{\Phi}_{S,S}\mathbf{\Phi}_{U,U}^{-1} + (1+\alpha)\mathbf{I}][\alpha\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{S,S} + (1-\alpha)\mathbf{I}]. \quad \text{(A1)}$$

Inverse matrix $\mathbf{\Phi}_{U,U}^{-1}$ can be decomposed by its eigenvectors $\mathbf{u}_i, i=1,n$

$$\mathbf{\Phi}_{U,U}^{-1} = \sum_{i=1}^{m} \frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^H + \sum_{i=m+1}^{n} \frac{1}{\sigma_N^2}\mathbf{u}_i\mathbf{u}_i^H, \quad \text{(A2)}$$

where $\mathbf{u}_i, i=1,m$, are eigenvectors of the interference signals subspace, and $\mathbf{u}_i, i=m+1,n$ are eigenvectors of the noise subspace; $\lambda_i, i=1,m$ are corresponding eigenvalues of the interference signals subspace, and $\sigma_N^2, \sigma_N^2=\Phi_{N,N}$ is common eigenvalue for eigenvectors of the noise subspace. Substituting (A2) and Eq. (4) into (A1) and taking into account $\mathbf{h}_1^H\mathbf{h}_1=1$ and $\sigma_N^2/\Phi_{s,s}\mathbf{I} \to \mathbf{0}$, ($\Phi_{s,s}$, is signal power), the auxiliary matrix $\mathbf{\Phi}$ can be approximated by

$$\mathbf{\Phi} \approx \alpha^2\Phi_{s,s}^2\gamma\mathbf{v}_p\mathbf{v}_p^H, \quad \text{(A3)}$$

where $\gamma$ is a real positive constant defined by $\gamma=\mathbf{v}_p^H\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\mathbf{v}_p$. From Eq. (A3) it is clear that the principal eigenvector of $\mathbf{\Phi}$ is approximately equal to $\mathbf{v}_p$, e.g., the principal eigenvector of $\hat{\mathbf{\Phi}}_{S,S}$.

## References and links

[1]R. A. Monzingo and T. W. Miller, *Introduction to Adaptive arrays*, (Wiley, New York, 1980).
[2]K. U. Simmer *et al.*, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. S. Brandstein and D. B. Ward (Springer, Berlin, 2001), Chap. 3, pp. 39–60.
[3]H. L. Van Trees, *Optimum Array Processing* (Wiley-Interscience, New York, 2002).
[4]J. E. Greenberg *et al.*, "Evaluation of an adaptive beamforming method for hearing aids," J. Acoust. Soc. Am. **91**, 1662–1676 (1992).
[5]O. Hoshuyama *et al.*, "A realtime robust adaptive microphone array controlled by a SNR estimate," *Proc. ICASSP98*, 3605–3608 (1998).

[6]Z. M. Saric and S. T. Jovicic, "Adaptive microphone array based on pause detection," ARLO **5**(2), 68–74 (2004).
[7]S. T. Jovicic *et al.*, "Application of the maximum signal to interference ratio criterion to the adaptive microphone array," ARLO **6**(4), 232–237 (2005).
[8]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**(4), 943–950 (1979).