Adaptive microphone array free of the desired speaker cancellation combined with postfilter

Slobodan Jovicic[a] and Zoran Saric[b]

[a]School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serby
[b]Institute for Experimental Phonetics and Speech Pathology, Gospodar Jovanova 35, 11000 Belgrade, Serby
jovicic@etf.bg.ac.yu

The optimal microphone array includes two processing blocks - minimum variance distortionless response (MVDR) beamformer and the single-channel Wiener filter, which acts as post-filter. The main drawback of 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, the estimation of the unknown desired speaker's transfer function was proposed. The estimation is based on the imperfect signal and the interference covariance matrices estimated from available microphone signals during speaker activity and pause of speech respectively. As MVDR beamformer suppresses coherent interference, post-filter has to reduce diffuse acoustic noise. The post-filter proposed in this paper is developed under assumption that complex coherence function is unknown but time invariant. The proposed *two step post-filter* (TS+post) algorithm was tested on simulated room with reverberation, and compared with some known post-processing algorithms with rather good results.

1. Introduction

Microphone arrays provide promising solution for the various applications such as video- conferences, hands-free telephony, hearing aids and speech recognition in noisy environment. They permit distant, hands-free signal acquisition by the directional discrimination, allowing for reduction of undesired noise sources.

The optimal microphone signals processing includes two blocks [1]. The first one is Minimum Variance Distortionless Response (MVDR) beamformer and the second one is single-channel Wiener filter, which acts as post-filter. In the real room environment there are both coherent and diffuse noises. MVDR beamformer attenuates mainly coherent noise, while post filter attenuates diffuse one.

The main problem in application of the MVDR beamformer in the room with reverberation is the cancellation of the desired speaker caused by the reflections from the walls [3], [11], [4]. The unknown desired speaker transfer function is usually approximated by direct path transfer function. The correlation of the direct path signal with reflections is the main reason for the signal cancellation [4]. There are many tools to improve performance of the MVDR beamformer. General approach is to make it robust by diagonal loading [2] or by some similar robustification techniques. Unfortunately, these methods also reduce capability of the beamformer to attenuate interferences. Other techniques use pauses in speech signal to estimate parameters of the beamformer [3], [11], [4]. These methods provide good interference attenuation without desired signal cancellation. Speech signal can be further improved by applying maximum signal to interference criterion instead of minimum variance one [13], [12]. This method calls for signal and interference matrices that are estimated in pause of speech and in pause of interference respectively. The estimation of the signal covariance matrix is difficult because interference is almost always present. The method proposed in [10] solves this problem.

The second, post-processing block was examined by Zelinski [5] who used the input channel auto- and cross-spectral densities to estimate a Wiener post-filter. The use of such a post-filter with the sub-array beamforming microphone array was thoroughly investigated by Marro *et al.* [6]. While the Zelinski post-filter shows reasonable performance, its formulation is based upon the unrealistic assumption of zero correlation of the channels noise in all frequency bins. To overcome this problem, I.A. McCowan *et al.* [7], used complex coherence for the ideal diffuse noise field to estimate post filter. This improves noise attenuation without additional signal distortion.

The alternative approach proposed in [14] is based on two assumptions: a) signal and noise are uncorrelated; b) complex coherence function is time invariant. The first assumption is also used in [5], [6], [7]. The second assumption is less restrictive then the assumptions used in Zelinski's and McCowan's algorithms. The algorithm from [14] provides high noise reduction without estimation of the input covariance matrix. In addition, it can be applied even the diffuse noise field assumption doesn't hold.

In this paper we combined beamformer proposed in [10] with post filter proposed in [14], as a new *two step post-filter* (TS+post) algorithm. This combined algorithm was tested on simulated room with reverberation, and compared with known microphone array post-processing algorithms.

2. MVDR beamformer in reverberant room

The optimal processing of the microphone signals in minimum mean squared error (MMSE) sense is depicted in Fig.1. The first processing block is MVDR beamformer.

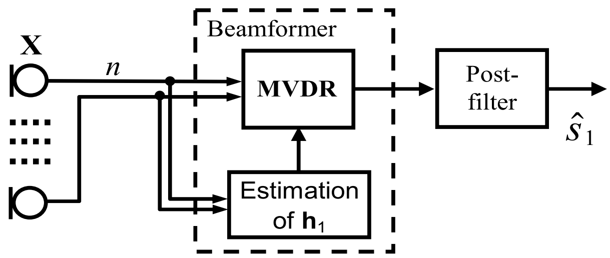

Fig.1 MVDR beamformer and post-filter.

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 DFT domain. All signals are represented by the complex DFT coefficients with central frequency f. For the sake of simplicity the frequency will be omitted, i.e. $x = x(f)$. Column vector $\mathbf{x}$ of the n microphone signals can be described 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 excitation of desired signal s_1, and n-column vector $\mathbf{h}_1$ is its transfer function containing both direct path and reflections.

The vector $\mathbf{U}$ is sum of responses to the interference signal vector $\mathbf{S}_I$, $\mathbf{S}_I = [s_2,\ldots,s_{m+1}]'$ and 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 x 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 $E\{.\}$ denotes the statistical expectation operator.

The output of the MVDR beamformer, 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 $\mathbf{\Phi}_{U,U} = E\{\mathbf{U}\mathbf{U}^H\}$ has to be estimated from available measurements $\mathbf{x}$ during absence of desired speech [4]. Desired signal transfer vector $\mathbf{h}_1$ includes both direct path and reflections from the walls. The problem is that transfer vector $\mathbf{h}_1$ is not a priori known in reverberant environment. The common used direct path transfer vector instead of $\mathbf{h}_1$ causes unwanted desired speech cancellation [3], [4]. To prevent this, the actual $\mathbf{h}_1$ has to be estimated.

Transfer vector $\mathbf{h}_1$ can be estimated from the imperfect estimate of the signal covariance matrix $\mathbf{\Phi}_{S,S} = E\{\mathbf{x}\mathbf{x}^H\}$ using principal eigenvector. The problem is that interference is almost always present and estimate of signal matrix is contaminated by interference one by:

$$\hat{\mathbf{\Phi}}_{S,S} = \gamma \mathbf{\Phi}_{S,S} + (1-\gamma)\mathbf{\Phi}_{U,U}, \quad 0.5<\alpha<1, \tag{4}$$

where γ is mixing scalar. The second term in Eq.(4) significantly degrades estimation of $\mathbf{h}_1$.

The improved estimate of $\mathbf{h}_1$ can be obtained under 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 desired signal power $\Phi_{N,N} << \Phi_{s,s}$.

Let us define auxiliary matrix $\mathbf{\Phi}_a$ as:

$$\mathbf{\Phi}_a = \hat{\mathbf{\Phi}}_{S,S}\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\hat{\mathbf{\Phi}}_{S,S}. \tag{5}$$

Under assumptions A1, A2, A3, the principal eigenvector $\mathbf{v}_p$ of $\mathbf{\Phi}_a$ can be used as approximation of the principal eigenvector of $\mathbf{\Phi}_{S,S}$. The proof is given in [10]. The estimate of unknown transfer function $\mathbf{h}_1$ is:

$$\hat{\mathbf{h}}_1 = C_\varphi \, \mathbf{v}_p, \tag{6}$$

where $\mathbf{v}_p$ is principal eigenvector of $\mathbf{\Phi}_a$, and where C_φ is a unit magnitude complex multiplier that compensate the signal delay [10]. The estimate of the weight vector of the MVDR beamformer is:

$$\mathbf{w}_{MVDR} = \frac{\hat{\mathbf{\Phi}}_{U,U}^{-1}\hat{\mathbf{h}}_1}{\hat{\mathbf{h}}_1^H \hat{\mathbf{\Phi}}_{U,U}^{-1}\hat{\mathbf{h}}_1}. \tag{7}$$

The usually used method to make MVDR beamformer robust against estimation errors of $\hat{\mathbf{\Phi}}_{U,U}$ and $\hat{\mathbf{h}}_1$ is to apply diagonal loading by:

$$\mathbf{w}_{DL} = \frac{\left(\hat{\mathbf{\Phi}}_{UU} + \delta\mathbf{I}\right)^{-1}\hat{\mathbf{h}}_1}{\hat{\mathbf{h}}_1^H \left(\hat{\mathbf{\Phi}}_{UU} + \delta\mathbf{I}\right)^{-1}\hat{\mathbf{h}}_1}, \tag{8}$$

where scalar constant δ makes compromise between robustness and high interference suppression. In the next, $\mathbf{w}_{DL}$ will be used instead of $\mathbf{w}_{MVDR}$.

3. Post-filter design

From the Wiener filtering theory, the optimal post-filter is:

$$H_{post} = \frac{\phi_{ss}^{out}}{\phi_{ss}^{out} + \phi_{uu}^{out}}, \tag{9}$$

where ϕ_{ss}^{out} is signal power, and ϕ_{uu}^{out} is noise power on output. The unknown signal and noise power have to be estimated from available signal measurements. Without loss of generality, we will assume:

$$\mathbf{h}_1^H \mathbf{h}_1 = m \ . \tag{10}$$

By taking into account Eq.(1) and Eq.(10), the average power of the microphone signals $\phi_{xx} = \frac{1}{m} E\{\mathbf{x}^H \mathbf{x}\}$ is:

$$\phi_{xx} = \frac{\mathbf{h}_1^H \mathbf{h}_1}{m}\phi_{ss} + \phi_{uu} = \phi_{ss} + \phi_{uu}, \tag{11}$$

where ϕ_{ss} and ϕ_{uu} are signal and interference power respectively. The output power ϕ_{yy} of the beamformer is:

$$\begin{aligned}\phi_{yy} &= \phi_{ss}\mathbf{w}_{DL}^H(\mathbf{h}_1\mathbf{h}_1^H)\mathbf{w}_{DL} + \mathbf{w}_{DL}^H \mathbf{\Phi}_{U,U}\mathbf{w}_{DL} = \\ &= \alpha\phi_{ss} + A_\Gamma \phi_{uu},\end{aligned} \tag{12}$$

where real scalar α is signal power attenuation factor that is approximately equal to one for $\hat{\mathbf{h}}_1 \approx \mathbf{h}_1$. The real scalar $A_\Gamma = \mathbf{w}_{DL}^H \mathbf{\Gamma}_{nn}\mathbf{w}_{DL}$ is noise power attenuation factor [1], where $\mathbf{\Gamma}_{uu} = \mathbf{\Phi}_{U,U}/\phi_{uu}$ is noise coherence matrix. If $\mathbf{\Gamma}_{uu}$ is time invariant then A_Γ is time invariant too. If we apply appropriate δ in Eq.(8) then $A_\Gamma < \alpha < 1$ and there is unique solution for linear system Eq.(11) and Eq.(12) by:

$$\phi_{ss} = \frac{\phi_{yy} - A_\Gamma\phi_{xx}}{\alpha - A_\Gamma}, \quad \phi_{uu} = \frac{\alpha\phi_{xx} - \phi_{yy}}{\alpha - A_\Gamma}. \tag{13}$$

The average microphone signals power ϕ_{xx} and beamformer output power ϕ_{yy} can be recursively estimated by [1]:

$$\hat{\phi}_{xx}(t) = \lambda\hat{\phi}_{xx}(t-1) + (1-\lambda)\mathbf{x}^H(t)\mathbf{x}(t)/m \tag{14}$$

$$\hat{\phi}_{yy}(t) = \lambda\hat{\phi}_{yy}(t-1) + (1-\lambda)\mathbf{w}_{DL}^H\mathbf{x}(t)\mathbf{x}^H(t)\mathbf{w}_{DL}, \tag{15}$$

where t denotes index of DFT data block. Positive scalar λ, $0<\lambda<1$ is exponential weighting factor. A_Γ is unknown and should be estimated too. Let us define the following auxiliary variable $\tilde{A}_\Gamma(t)$ by:

$$\widetilde{A}_\Gamma(t) = \frac{\hat{\phi}_{yy}(t)}{\hat{\phi}_{xx}(t)}. \tag{16}$$

Expectation value of $\widetilde{A}_\Gamma(t)$ is:

$$E\{\widetilde{A}_\Gamma(t)\} = \frac{\alpha\phi_{ss} + A_\Gamma\phi_{nn}}{\phi_{ss} + \phi_{nn}} \tag{17}$$

During speech interval ($\phi_{ss} >> \phi_{nn}$), the auxiliary variable $\widetilde{A}_\Gamma(t)$ is approximately equal to α. Contrary, during pause of speech ($\phi_{ss} << \phi_{nn}$), the auxiliary variable $\widetilde{A}_\Gamma(t)$ is approximately equal to A_Γ. Hence, A_Γ can be estimated by recursive averaging of $\widetilde{A}_\Gamma(t)$ by the first order IIR filter with different constants λ_p and λ_n for the positive and negative slope of $\hat{A}_\Gamma$ by:

$$\hat{A}_\Gamma(t) = \begin{cases} \lambda_n \hat{A}_\Gamma(t-1) + (1-\lambda_n)\widetilde{A}_\Gamma(t), & for\ \widetilde{A}_\Gamma(t) < \hat{A}_\Gamma(t-1) \\ \lambda_p \hat{A}_\Gamma(t-1) + (1-\lambda_p)\widetilde{A}_\Gamma(t), & for\ \widetilde{A}_\Gamma(t) \geq \hat{A}_\Gamma(t-1) \end{cases},$$

$$0 < \lambda_n < \lambda_p < 1, \tag{18}$$

By replacing the true values of ϕ_{xx}, ϕ_{yy} and A_Γ with their estimates into Eq.(13), and taking into account $\hat{\phi}_{uu}^{out}(t) = \hat{A}_\Gamma(t)\hat{\phi}_{uu}(t)$ and $\phi_{ss}^{out} = \alpha\phi_{ss}$, the post-filter is:

$$\hat{H}_{post}(t) = \frac{\alpha\hat{\phi}_{ss}(t)}{\alpha\hat{\phi}_{ss}(t) + \hat{A}_\Gamma(t)\hat{\phi}_{uu}(t)} = \frac{\alpha(\hat{\phi}_{yy}(t) - \hat{A}_\Gamma(t)\hat{\phi}_{xx}(t))}{(\alpha - \hat{A}_\Gamma(t))\hat{\phi}_{yy}(t)} \tag{19}$$

For appropriate diagonal loading, scalar α is close to 1 and:

$$\hat{H}_{post}(t) \approx \frac{\hat{\phi}_{yy}(t) - \hat{A}_\Gamma(t)\hat{\phi}_{xx}(t)}{(1 - \hat{A}_\Gamma(t))\hat{\phi}_{yy}(t)}. \tag{20}$$

To reduce the estimation errors an additional constraint:

$$0 \leq \hat{H}_{post}(t) \leq 1, \tag{21}$$

has to be applied.

4. Experimental results

The proposed combined MVDR and post-processing algorithm has been examined in a room with reverberation simulated by Allen's image method [9]. The microphone array consisted of 5 microphones with equidistant spacing of 10cm. There was two acoustic sources: 1) desired speaker 50cm in front of the microphone array, and 2) Gaussian noise source at 2m distance at angle of 45^0 (Fig.2). Critical distance boundary was calculated from the room model at which the direct path power is equal to the reverberant power. Sampling frequency was 10 kHz.

The quality of the restored speech was evaluated by two distance measures: (a) segmental signal to interference ratio enhancement (SNRE), and (b) log-area-ratio distance measure (LAR). The following algorithms are tested: 1) Zelinski [5], 2) APAB [1], 3) McCowan [7], 4) Saric [14], and 5) proposed (TS+post) algorithm.

Simulation system generates three sets of microphone signals [1]: (a) a room's response to the desired signal $\mathbf{x}_s$, $\mathbf{x}_s = \mathbf{h}_1 s$, where $\mathbf{h}_1$ is room impulse response to the speech excitation, (b) a room's response to the noise excitation $\mathbf{x}_v$, $\mathbf{x}_v = \mathbf{h}_2 v$, where v is Gaussian noise and $\mathbf{h}_2$ is room impulse response to noise source, and (c) a vector of test microphone signals $\mathbf{x} = \mathbf{x}_s + \beta\mathbf{x}_v$ with an input signal to noise ratio (SNR) controlled by scale factor β. The vector **x** is input to master algorithm. The signals $\mathbf{x}_s$ and $\mathbf{x}_v$ are inputs to two slave algorithms that use copies of master post-filter parameters. The outputs of master and two slave algorithms are z_{s+v}, z_s and z_v respectively [1]. ($z = \hat{H}_{post}\mathbf{w}_{DL}^H\mathbf{x}$, where $\mathbf{w}_{DL}$ is weight vector of the MVDR beamformer).

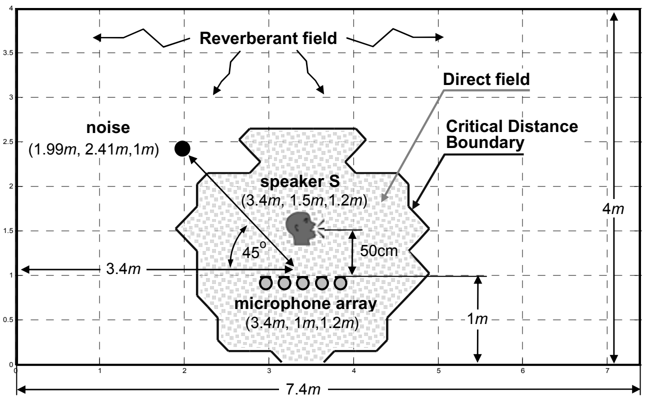

Fig.2 Simulated room with reverberation T_{60} = 300ms.

Segmental signal-to-noise ration enhancement (SNRE) in term of input SNR is displayed in Fig.3. Post-filter algorithms McCowan [7] and Saric have similar SNRE, much better then Zelinski and APAB algorithms. Combined (TS+post) outperforms all other algorithms, especially at high input SNR.

Speech degradation is measured with LAR distance measure and displayed in Fig.4. The proposed (TS+post) algorithm also outperforms all other tested algorithms.

5. Conclusions

The combined microphone array algorithm with MVDR beamformer as a first processing step and a single channel post-filter as a second processing step is proposed in this paper. The proposed algorithm is capable to suppress combination of the coherent and diffuse noise in reverberant room.

The first processing block is MVDR beamformer with estimation of the desired signal transfer function when lot of reflections are present. The use of the estimated transfer function prevents desired signal cancellation and provides high attenuation of the coherent interference. Estimation of the unknown signal transfer function is based on unperfected estimate of the signal and interference matrix. If the signal-to-white noise ratio is sufficiently high the estimate of the signal transfer function is insensitive to the contamination of the signal covariance matrix with coherent interference.

The proposed post-filter is based on estimation of the noise power attenuation factor. This factor depends on the noise coherence matrix and it is time invariant if the sources and microphone array don't change their position. The proposed post filter algorithm doesn't use assumption of the ideal diffuse noise field as Zelinski and McCowan algorithms do.

Its performance doesn't fall down if the ideal diffuse noise field doesn't hold.

In real reverberant room, interference acts both as coherent and diffuse noise depending of the distance from the microphone array. If the interference is inside the critical distance boundary, it acts mainly as coherent interference. Otherwise, it acts mainly as diffuse noise source. The proposed combined MVDR-postfilter algorithm provides optimal signal enhancement in MMSE sense in the situation when both diffuse and coherent are present. That is the reason for so god performance of the proposed algorithm.

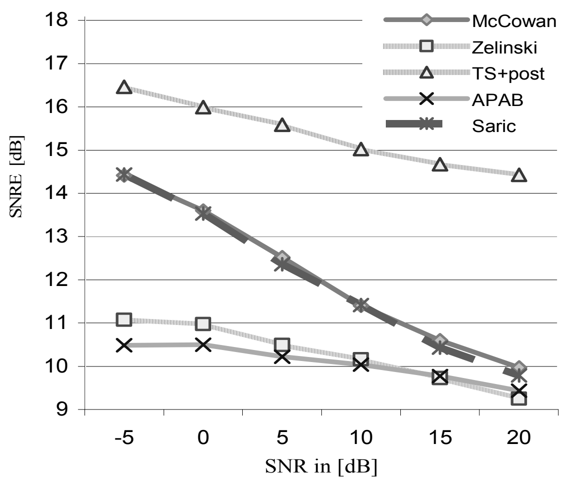

Fig.3 Segmental signal-to-noise enhancement (SNRE) in term of input SNR.

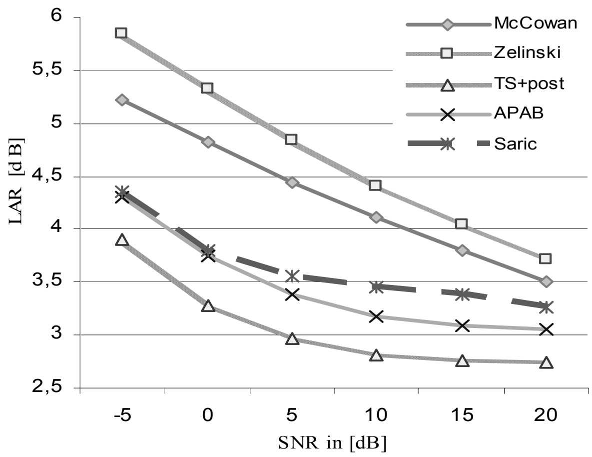

Fig.4 LAR distortion measure in term of the input SNR.

Acknowledgments

This work was supported by Ministry of Science of the Republic Serbia under Grant number 148028G.

References

[1] K. U. Simmer, J. Bitzer, C. Marro, "Post-filtering techniques", In M. Brandstein, D. Ward (eds), *Microphone Arrays,* New York: Springer, ch. 3, 36–60 (2001).

[2] H. L. Van Trees, *Optimum Array processing,* New York: John Wiley & Sons, Inc. (2002).

[3] J. E. Greenberg, P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids", *J. Acoust. Soc. Amer.* 91(3), 1662-1676 (1992).

[4] Z. M. Saric, S. T. Jovicic, "Adaptive microphone array based on pause detection," *Acoustics Research Letters Online (ARLO)* 5(2), 68-74 (2004).

[5] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms", *Proc. ICASSP88*, 2578–2581 (1988).

[6] C. Marro, Y. Mahieux, K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering", *IEEE Trans. Speech Audio Process*. 6(3), 240–259 (1998).

[7] I. A. McCowan, H. Bourlard, "Microphone array post-filter for diffuse noise field", *Proc. ICASSP-02*, 905-908 (2002).

[9] J. B. Allen, D. A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Amer.* 65(4), 943-950 (1979).

[10] I. Papp, Z. Šarić, S. Jovičić, N. Teslić, "Adaptive microphone array for unknown desired speaker's transfer function", *JASA Express Letters*. 122(2), EL44-49 (2007).

[11] O. Hoshuyama, B. Begasse, A. Sugiyama, A. Hirano, "A realtime robust adaptive microphone array controlled by an SNR estimate", *Proc. ICASSP98*, 3605-3608 (1998).

[12] S. T. Jovicic, Z. M. Saric, S. R. Turajlic, "Application of the maximum signal to interference ratio criterion to the adaptive microphone array", *Acoustics Research Letters Online (ARLO)* 6(4), 232-237, (2005).

[13] D. R. Morgan, "Adaptive algorithms for solving generalized eigenvalue signal enhancement problem", *Signal Processing* 84, 957-968 (2004).

[14] Z. Saric, S. Jovicic, M. Janev, I. Paap, Z. Marceta; "Microphone Array Post-filter Based on Noise Power Attenuation Factor and A Priori Knowledge of Noise Field Coherence", *Proc. of International Conference SPECOM 2007*, Moscow, Russia, 252-258 (2007).