Archives of Acoustics Vol. **46**, No. 4, pp. 611–621 (2021), doi: 10.24425/aoa.2021.138154

Research Paper

Performance Analysis of MVDR Beamformer Applied on an End-fire Microphone Array Composed of Unidirectional Microphones

Zoran ŠARIĆ(1)*, Miško SUBOTIĆ(1), Ružica BILIBAJKIĆ(1),
Marko BARJAKTAROVIĆ(2), Nebojša ZDRAVKOVIĆ(3)

(1) *Laboratory of Acoustics, Life Activities Advancement Center*
Serbia; e-mail: {m.subotic, r.bilibajkic}@add-for-life.com
*Corresponding Author e-mail: sariczoran@yahoo.com
Corresponding Author ORCID: 0000-0001-9964-9974

(2) *Faculty of Electrical Engineering, University of Belgrade*
Serbia; e-mail: mbarjaktarovic@etf.bg.ac.rs

(3) *Faculty of Medical Sciences, University of Kragujevac*
Serbia; e-mail: nzdravkovic@medf.kg.ac.rs

(*received October 23, 2020; accepted September 8, 2021*)

Microphone array with minimum variance (MVDR) beamformer is a commonly used method for ambient noise suppression. Unfortunately, the performance of the MVDR beamformer is poor in a real reverberant room due to multipath wave propagation. To overcome this problem, we propose three improvements. Firstly, we propose end-fire microphone array that has been shown to have a better directivity index than the corresponding broadside microphone array. Secondly, we propose the use of unidirectional microphones instead of omnidirectional ones. Thirdly, we propose an adaptation of its adaptive algorithm during the pause of speech, which improves its robustness against the room reverberation and deviation from the optimal receiving direction. The performance of the proposed microphone array was theoretically analyzed using a diffuse noise model. Simulation analysis was performed for combined diffuse and coherent noise using the image model of the reverberant room. Real room tests were conducted using a four-microphone array placed in a small office room. The theoretical analysis and the real room tests showed that the proposed solution considerably improves speech quality.

Keywords: adaptive beamforming; ambient noise suppression; differential microphone array; end-fire microphone array, MVDR beamformer.

1. Introduction

In real speech communication, ambient noise significantly degrades the quality of speech. An effective method for ambient noise suppression is to use a microphone array in combination with proper multichannel signal processing. A commonly used method for processing microphone array signals is adaptive beamforming, which typical representative is minimum variance distortionless response (MVDR) beamformer (Capon, 1969). Parameters of the MVDR beamformer are estimated to minimize the power of the ambient noise with unit gain constraint for the desired speech signal. A more general linearly constrained minimum variance (LCMV) beamformer (Frost, 1972), uses multiple constraints to enhance some desirable features of the beamformer.

The performance of the MVDR beamformer is poor in a reverberant room environment. In addition, the presence of the calibration and the steering errors cause unwanted suppression and degradation of the desired speech signal. This problem can be overcome by diagonal loading (Van Trees, 2004), using some robust estimation method, or by adaptation in the pause of

speech, (Hoshuyama *et al.*, 1999; Saric, Jovicic, 2004; Jovicic *et al.*, 2005; Wölfel, McDonough, 2009). Analysis of the MVDR beamformer and its improvements are presented in (Pan *et al.*, 2014; 2015; Benesty *et al.*, 2008; Wang *et al.*, 2019).

Some other noise suppression methods use a superdirective beamformer with a fixed beam pattern optimized for a particular noise type (Elko, 2004; Bitzer, Simmer, 2001). The advantage of the superdirective beamformers is their low sensitivity to moderate steering errors. Their drawback is their sensitivity to the white noise. Moreover, they are optimized for a particular noise type which makes them less efficient for suppression of the other noise types[1].

Another method for multichannel noise reduction uses generalized singular value decomposition (GSVD) that minimizes the mean squares error between desired speech signal and the filtered received microphone signals (Spriet *et al.*, 2002). This method is suitable for the small microphone array used in hearing aid devices. A similar method (Papp *et al.*, 2007), provides robust adaptive beamforming by using signal covariance matrix, noise covariance matrix and the principal eigenvector of an auxiliary matrix calculated from the signal and the noise matrices.

A microphone-array based system for full-duplex hands-free voice communication integrated with TV technology was proposed in (Papp *et al.*, 2011). The system provides a comfortable conversation and includes superdirective beamformer steered by direction-finding module, postprocessing module, acoustic echo canceller, stationary noise reduction module and automatic gain control.

In the solution (Šarić *et al.*, 2019), the authors considered a teleconference scenario where two speakers share the same microphone array. Speakers' activity is detected by a control module using signals from four superdirective beamformers with different beam patterns. In accordance with the speakers' activity, the control module switches on one of the two MVDR beamformers directed toward each of the speakers.

Post-processing by the time-varying Wiener filter estimated from the coherence matrix is widely used method in teleconference applications (Zelinski, 1988; Marro *et al.*, 1998; Simmer *et al.*, 2001; McCowan, Bourlard, 2003; Saric *et al.*, 2011). Post-processing efficiently reduces diffuse noise and reverberation of the room, but it also changes power spectral density (PSD) of the speech which is unacceptable for the applications that demand high-quality speech recordings.

Blind source separation (BSS) methods separate the desired speech from ambient noise in order to enhance the speech signal. The most of the BSS methods are based on second-order statistics in which nonstationarity of the speech signal is used to optimize parameters of the spatial filter (Parra, Spence, 2000; Parra, Alvino, 2002; Wang *et al.*, 2010). The serious problem in the application of the BSS method is source permutation which may cause degradation of speech PSD.

One class of BSS methods, called degenerate unmixing estimation technique (DUET) (Yilmaz, Rickard, 2004), uses the assumption that competitive signals are nonoverlapping in time and frequency domain. The advantage of DUET is that it can even handle cases when there are more signals than microphones. It uses instantaneous time delay estimates and the estimates of signal attenuation in each frequency bin to calculate the binary mask used to separate the desired signal from other interference signals. DUET can efficiently enhance the speech in a room with moderate reverberation time. Its disadvantage is the degradation of speech due to errors in the evaluation of the binary mask caused by noise and reverberation of the room.

Some of the modern noise suppression methods use a spherical microphone array composed of many microphones placed on the rigid sphere (McDonough, Kumatani, 2012). Although the spherical microphone array outperforms the linear microphone array with similar dimensions, it is a rather expensive solution because it requires a specially designed spherical microphone array with 32 to 64 microphones. Besides this, it requires special multichannel analog-to-digital converter (ADC) for the acquisition of a large number of independent signals, and it needs a powerful computation platform to process a large number of these signals in real time.

In recent studies (Krecichwost, *et al.*, 2019; 2020) a microphone arrays are used in multichannel acquisition devices for computer-aided diagnostics of speech disorders. A specially designed head-worn acoustic mask with spatially arranged microphones located in front of subject's mouth enables recording of the speech signal with increased signal-to-noise ratio of the weak speech components.

Computational auditory scene analysis (CASA) is a method based on perceptual principles of auditory scene analysis (Wang, Brown, 2006). In CASA, each time-frequency (T-F) element is classified as speech-dominant or noise-dominant. An ideal binary mask (IBM) applied to noisy speech signal separates speech from noise. Estimation of the IBM is a binary classification problem where supervised learning is employed to predict the label of each T-F unit (Wang, Chen, 2018). The binary mask can be estimated from a single microphone or multiple microphones (binaural or multichannel processing). The CASA approach is suitable for various noise types where improves speech intelligibility and reduces the word error rate of automatic speech recognition (ASR) even for low speech-to-noise

[1]Superdirective beamformer is usually optimized for spherical or cylindrical diffuse noise field.

ratio (Chen, *et al.*, 2014). On the other hand, according to our knowledge, there is no report about speech distortion when it is applied to noise-free speech signal.

In accordance with the state of the art in the field of speech enhancement and taking into account that some applications demand speech enhancement without speech distortion, the aim of this paper is to propose a low-cost noise reduction method without speech distortion.

One of the approaches that provides suppression of the ambient noise without any speech degradation is the MVDR beamformer. As the performance of the MVDR beamformer is poor in a real reverberant room, we propose three improvements of the basic method. Firstly, we propose an end-fire linear microphone array which is proved to have better directivity index (DI) than corresponding broadside microphone array (Pan *et al.*, 2014; Trucco *et al.*, 2015; Soede *et al.*, 1993; Kates, Weiss, 1996; Greenberg, Zurek, 2001).

Secondly, to further increase DI, we propose using unidirectional instead of omnidirectional microphone capsules. Although it is intuitively clear that the use of unidirectional microphones improves DI compared to the use of omnidirectional microphones, to the authors' knowledge, no theoretical analysis has been performed so far to confirm this fact and to quantitatively evaluate signal to noise ratio (SNR) gain.

Thirdly, we propose an adaptation of the MVDR beamformer during pause of speech, which prevents desired speech cancellation and improves robustness against deviation of the speaker's position from optimal direction (Saric, Jovicic, 2004; Jovicic *et al.*, 2005). Although each of these methods individually improves SNR, the best performance is obtained by applying these methods together.

The paper is organized as follows.

In Sec. 2, we theoretically analyzed the performance of the end-fire microphone array with an arbitrary number of unidirectional and omnidirectional microphones. In Subsec. 2.1, we defined a signal model for the reverberant room environment that will be used in further analysis. Elements of the LCMV beamformer and some common notations are presented in Subsec. 2.2. The model of the unidirectional microphone is presented in Subsec. 2.3. The model of the array composed of an arbitrary number of unidirectional and omnidirectional microphones and its application to the LCMV beamformer is presented in Subsec. 2.4. Theoretical analysis of the SNR gain of the microphone array in the diffuse noise field is presented in Subsec. 2.5. White noise gain (WNG) for the proposed microphone array is analyzed in Subsec. 2.6.

Performance analysis of the proposed microphone array is presented in Sec. 3. The SNR gain for three configurations of the microphone array with limited WNG is evaluated for the diffuse noise field in Subsec. 3.1. The performance of the proposed microphone array in the simulated room environment is presented in Subsec. 3.2, while the performance of the laboratory model of the proposed microphone array in the real reverberant room is presented in Subsec. 3.3. Results are discussed in Sec. 4. Conclusions are presented in Sec. 5.

2. Model of the proposed microphone array (Method)

2.1. Signal model

Let us consider a microphone array composed of M microphones, as shown in Fig. 1. In accordance with the reverberant room model (Benesty, *et al.*, 2008) measurements $y_m(t)$ at m-th microphone are

$$y_m(t) = \mathbf{g}_m^{\mathrm{T}}\mathbf{s}(t) + v_m(t), \tag{1}$$

where $\mathbf{s}(t) = [s(t), ..., s(t - L_g + 1)]^{\mathrm{T}}$ is a column vector of last L_g samples of zero mean speech signal, $\mathbf{g}_m = [g_{m,1}, ..., g_{m,L_g}]^{\mathrm{T}}$ is L_g-column vector that represents room impulse response from the desired speaker to the m–th microphone, $v_m(t)$ is additive noise uncorrelated with $s(t)$. Superscript T is transpose operator. The corresponding model in discrete Fourier transform (DFT) is

$$Y_m(\omega, l) = g_m(\omega)S(\omega, l) + V_m(\omega, l), \tag{2}$$

where $Y_m(\omega, l)$, $S(\omega, l)$, and $V_m(\omega, l)$ are DFT coefficients of $y_m(l)$, $s(l)$, and $v_m(l)$ respectively at segment l and angular frequency ω. Discrete Fourier transforms are performed on overlapping segments. For simplicity, the index l is omitted in the rest of the paper. Matrix form of the signal model (2) is

$$\mathbf{Y}(\omega) = \mathbf{G}(\omega)S(\omega) + \mathbf{V}(\omega), \tag{3}$$

where $\mathbf{Y}(\omega) = [Y_1(\omega)\ \ Y_2(\omega)\ \ Y_3(\omega)\ \ ...\ \ Y_M(\omega)]^{\mathrm{T}}$ is the M-column vector of complex microphone signals, $\mathbf{G}(\omega) = [g_1(\omega), ..., g_M(\omega)]^{\mathrm{T}}$ is the M-column transfer vector, and $\mathbf{V}(\omega) = [V_1(\omega), ..., V_M(\omega)]^{\mathrm{T}}$ is the M-column vector of the ambient noise. The output of the beamformer is the weighted sum of the microphone signals,

$$Z(\omega) = \mathbf{h}^H(\omega)\mathbf{Y}(\omega), \tag{4}$$

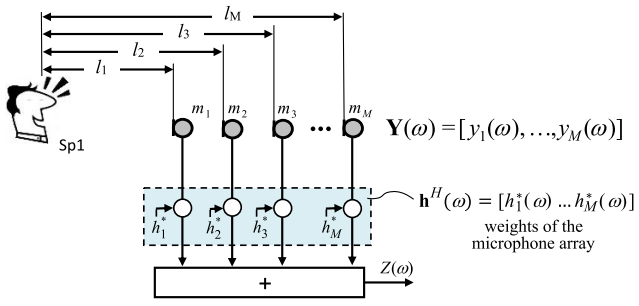

Fig. 1. Beamformer in DFT domain.

where $\mathbf{h}(\omega) = [h_1^*(\omega) \;\; h_2^*(\omega) \;\; h_3^*(\omega) \;\; \ldots \;\; h_M^*(\omega)]^H$ is the M-column vector of complex weights. Superscript H denotes complex conjugate transpose while * denotes conjugate operation. The commonly used methods for weights estimation are MVDR or LCMV estimators. In the following, we will focus on the LCMV estimator as a more general estimation method.

2.2. LCMV beamformer

Estimation of the LCMV beamformer's weights can be defined as conditional minimization (Frost, 1972)

$$\begin{aligned} \mathbf{h}_{\mathrm{LCMV}}(\omega) &= \arg \min_{\mathbf{h}} \mathbf{h}^H \mathbf{\Phi}_v(\omega)\mathbf{h} \\ \text{subject to} \quad & \mathbf{C}^H(\omega)\mathbf{h}(\omega) = \mathbf{f}, \end{aligned} \tag{5}$$

where $\mathbf{\Phi}_v(\omega) = E\left\{\mathbf{v}(\omega)\mathbf{v}^H(\omega)\right\}$ is the noise covariance matrix, $E\{\cdot\}$ is the mathematical expectation operator, $\mathbf{C}(\omega)$ is constraint matrix, and $\mathbf{f}$ is the desired response vector. Each constraint is defined by a column of the constraint matrix $\mathbf{C}(\omega)$ and the corresponding element of the desired response vector $\mathbf{f}$. Usually used constraint is unit gain for the desired speech defined by

$$\mathbf{G}^H(\omega)\mathbf{h}(\omega) = 1. \tag{6}$$

The solution of the optimization problem (6) is (Frost, 1972; DeFatta, 1988)

$$\mathbf{h}_{\mathrm{LCMV}}(\omega) = \mathbf{\Phi}_v^{-1}(\omega)\mathbf{C}(\omega)\left[\mathbf{C}^H(\omega)\mathbf{\Phi}_v^{-1}(\omega)\mathbf{C}(\omega)\right]^{-1}\mathbf{f}. \tag{7}$$

To ensure inversion of the matrix $\mathbf{\Phi}_v^{-1}(\omega)$, as well as to control white noise gain (WNG) of the beamformer, the diagonal loading has to be applied by (Van Trees, 2004)

$$\mathbf{\Phi}_v(\omega) = \widetilde{\mathbf{\Phi}}_v(\omega) + \delta\mathbf{I}, \tag{8}$$

where $\widetilde{\mathbf{\Phi}}_v(\omega)$ is noise covariance matrix estimated in the pause of speech, $\mathbf{I}$ is the unit matrix, and δ is a small positive scalar.

In this paper, the first constraint is the unit gain for the desired speech signal. The additional constraints are used to model unidirectional microphones.

2.3. Model of the unidirectional microphone

A unidirectional microphone can be modeled as a differential microphone composed of microphones m_k and m_{k+1}, Fig. 2, (Elko, 2004). Beam-pattern of the unidirectional microphone model is determined by the null steering angle θ_0 which is controlled by the distance between microphones d_c and the time delay τ_c

$$\theta_0 = \cos^{-1}(-c\tau_c/d_c), \tag{9}$$

where c is the speed of sound. Angles $\theta_0 = 90°$, $\theta_0 = 109°$, $\theta_0 = 125°$, and $\theta_0 = 180°$, define dipole, hypercardioid, supercardioid, and cardioid pattern, respectively (Elko, 2004). Without loss of generality, we consider the cardioid pattern, ($\theta_0 = 180°$) for which $\tau_c = d_c/c$. We used the cardioid pattern because it is widely used for low-cost electret microphone capsules.

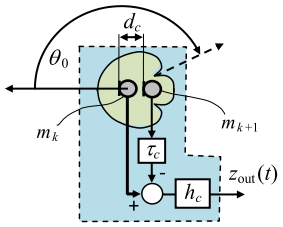

Fig. 2. Model of unidirectional microphone.

The flat frequency response of the model is obtained by the compensation factor $h_c(\omega)$:

$$h_c(\omega) = \frac{1}{1 - \exp(-j2\omega\tau_c)}. \tag{10}$$

2.4. Constraint-based unidirectional and omnidirectional microphone array model

Although it is obvious that unidirectional microphones built into an array of microphones increase its directivity, a quantitative analysis of that contribution has not been performed so far. In this paper, we perform this analysis using a model of microphone array in which each unidirectional microphone is represented by one constraint in constraint matrix $\mathbf{C}(\omega)$. Figure 3 shows a microphone array in which two omnidirectional microphones at positions $k(u)$ and $k(u)+1$ act as a differential microphone. The equivalent complex weights $h_{k(u)}(\omega)$ and $h_{k(u)+1}(\omega)$ of the microphones $m_{k(u)}$ and $m_{k(u)+1}$ are

$$\begin{aligned} h^*_{k(u)}(\omega) &= h^*_{k(u),\,k(u)+1}(\omega)h_c, \\ h^*_{k(u)+1}(\omega) &= -h^*_{k(u),\,k(u)+1}(\omega)h_c \exp\left(-j\omega\tau_c\right), \end{aligned} \tag{11}$$

where $h^*_{k(u),\,k(u)+1}(\omega)$ is complex weigh of u-th unidirectional model.

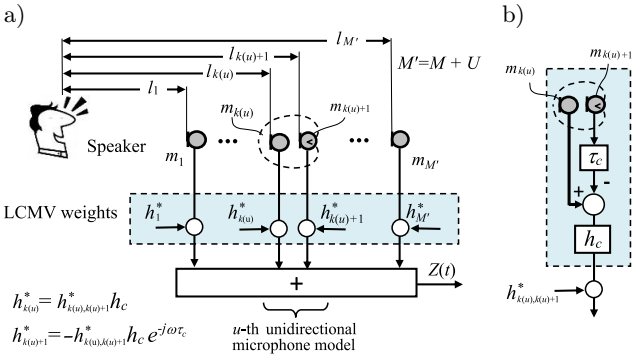

Fig. 3. (a) Microphone array with model of differential microphone at $k(u)$, $k(u)+1$ position; (b) unidirectional microphone model realized by microphones $m_k(u)$ and $m_k(u+1)$ with common weight $h_{k(u),\,k(u)+1}$.

Equations $(11)_1$ and $(11)_2$ define a linear constraint of the LCMV beamformer in a form

$$\mathbf{c}_u^H(\omega)\mathbf{h}(\omega) = 0, \tag{12}$$

where

$$\mathbf{c}_u^H(\omega) = [\overbrace{0 \ \ldots \ 0 \ \underbrace{\exp(j\omega\tau_c)}_{k(u)\text{ position}} \ \underbrace{1}_{k(u)+1} \ 0 \ \ldots \ 0}^{M\ (M\text{ is number of array elements})}], \tag{13}$$

The first column of the constraint matrix $\mathbf{C}(\omega)$ is the unit gain constraint vector, Eq. (6). Other M_u columns of $\mathbf{C}(\omega)$ are constraint vectors $\mathbf{c}_u = 1, ..., M_u$ defined by Eq. (13), where M_u is number of unidirectional microphone of the array. Finally, constraint matrix $\mathbf{C}(\omega)$ is

$$\mathbf{C}(\omega) = [\mathbf{G}(\omega) \ \ \mathbf{c}_1 \ \ \ldots \ \ \mathbf{c}_{M_u}]. \tag{14}$$

According to Eqs (6) and (12), the desired response vector is

$$\mathbf{f} = [\underbrace{1 \ \ 0 \ \ \ldots \ \ 0}_{M_u+1}]^{\mathrm{T}}. \tag{15}$$

Constraint matrix $\mathbf{C}(\omega)$ and desired response vector $\mathbf{f}$ have to be substituted into Eq. (7) to obtain LCMV beamformer's weights.

2.5. SNR gain of the microphone array model

The commonly used measure for signal enhancement by LCMV beamformer is signal to noise ratio (SNR) gain. Input SNR, denoted as $i\mathrm{SNR}(\omega)$, is the ratio of the input speech power, $\Phi_s(\omega) = E\left\{|S(\omega)|^2\right\}$, and noise power on reference microphone m_1, $\Phi_{v1}(\omega) = E\left\{|V_1(\omega)|^2\right\}$ (Pan *et al.*, 2014),

$$i\mathrm{SNR}(\omega) = \Phi_s(\omega)/\Phi_{v1}(\omega). \tag{16}$$

Speech power at the output of the beamformer is

$$P_{s\,\mathrm{out}}\left(\mathbf{h}_{\mathrm{LCMV}}(\omega)\right) = \Phi_s(\omega)\mathbf{h}_{\mathrm{LCMV}}^H(\omega)\mathbf{G}(\omega) \cdot \mathbf{G}^H(\omega)\mathbf{h}_{\mathrm{LCMV}}(\omega). \tag{17}$$

Noise power at the output of the beamformer $P_{v\,\mathrm{out}}\left(\mathbf{h}_{\mathrm{LCMV}}(\omega)\right)$ is

$$P_{v\,\mathrm{out}}\left(\mathbf{h}_{\mathrm{LCMV}}(\omega)\right) = \mathbf{h}_{\mathrm{LCMV}}^H(\omega)\mathbf{\Phi}_v(\omega)\mathbf{h}_{\mathrm{LCMV}}. \tag{18}$$

Taking into account Eqs (17) and (18), output SNR is

$$o\mathrm{SNR}\left(\mathbf{h}_{\mathrm{LCMV}}(\omega)\right) = \frac{P_{s\,\mathrm{out}}\left(\mathbf{h}_{\mathrm{LCMV}}(\omega)\right)}{P_{v\,\mathrm{out}}\left(\mathbf{h}_{\mathrm{LCMV}}(\omega)\right)} = \frac{\Phi_s(\omega)\mathbf{h}_{\mathrm{LCMV}}^H(\omega)\mathbf{G}(\omega)\mathbf{G}^H(\omega)\mathbf{h}_{\mathrm{LCMV}}(\omega)}{\mathbf{h}_{\mathrm{LCMV}}^H(\omega)\mathbf{\Phi}_v(\omega)\mathbf{h}_{\mathrm{LCMV}}} = i\mathrm{SNR}(\omega)\frac{\mathbf{h}_{\mathrm{LCMV}}^H(\omega)\mathbf{G}(\omega)\mathbf{G}^H(\omega)\mathbf{h}_{\mathrm{LCMV}}(\omega)}{\mathbf{h}_{\mathrm{LCMV}}^H(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{h}_{\mathrm{LCMV}}(\omega)}, \tag{19}$$

where $\mathbf{\Gamma}_v(\omega)$ is noise pseudo-coherence defined by

$$\mathbf{\Gamma}_v(\omega) = \mathbf{\Phi}_v(\omega)/\Phi_{v1}(\omega). \tag{20}$$

SNR gain of the LCMV beamformer is the ratio between output SNR, (19), and input SNR. Taking into account Eq. (19) and constraint (6), the SNR gain is

$$\mathrm{SNR}_{\mathrm{gain}}\left(\mathbf{h}_{\mathrm{LCMV}}(\omega)\right) = \frac{1}{\mathbf{h}_{LCMV}^H(\omega)\mathbf{\Gamma}_v(\omega)\mathbf{h}_{\mathrm{LCMV}}(\omega)}. \tag{21}$$

Substituting Eqs (20) and (7) into Eq. (21), SNR gain for the LCMV beamformer is

$$\mathrm{SNR}_{\mathrm{gain}}(\omega) = \frac{1}{\mathbf{f}^H\left[\mathbf{C}(\omega)^H\mathbf{\Gamma}_v^{-1}(\omega)\mathbf{C}(\omega)\right]^{-1}\mathbf{f}}. \tag{22}$$

From Eq. (22) we see that SNR gain depends on the noise pseudo-coherence matrix $\mathbf{\Gamma}_v(\omega)$, constraint matrix $\mathbf{C}(\omega)$, and desired response vector $\mathbf{f}$. The pseudo-coherence matrix $\mathbf{\Gamma}_v(\omega)$ depends on the spatial distribution of the noise power, while $\mathbf{C}(\omega)$ and $\mathbf{f}$ are fixed for the particular array structure.

2.6. White noise gain of the array

The white noise gain (WNG) shows the ability of the array to suppress spatially uncorrelated noise mostly caused by self-noise of the microphones. Omnidirectional and unidirectional microphones contribute differently to output noise power. Under assumption that all omnidirectional microphones have the same noise variance equal $\sigma_o^2(\omega) = E\left\{|\xi_m(\omega)|^2\right\}$, their contribution to the output noise power is (Pan *et al.*, 2014)

$$P_{n\_\mathrm{omni}}(\omega) = \sigma_o^2(\omega)\sum_{m\in O}|h_m(\omega)|^2, \tag{23}$$

where O is the set of omnidirectional microphones of the microphone array. Using Eqs $(11)_1$ and $(11)_2$, contribution of unidirectional microphones to the output noise power is

$$P_{n\_\mathrm{uni}}(\omega) = \sigma_U^2(\omega)\sum_{u\in U}\left(\left|h_{k(u)}(\omega)\right|^2 + \left|h_{k(u)+1}(\omega)\right|^2\right) = \sigma_{eU}^2(\omega)\sum_{u\in U}\left|h_{k(u),k(u)+1}(\omega)\right|^2, \tag{24}$$

where $\sigma_U^2(\omega)$ is white noise variance of microphones $m_k(u)$ and $m_k(u+1)$, $\sigma_{eU}^2(\omega) = 2\sigma_U^2(\omega)\left|h_c(\omega)\right|^2$ is variance of equivalent noise of the u-th unidirectional microphone. Assuming that omnidirectional and unidirectional microphones have the same self-noise, i.e. $\sigma_{eU}^2(\omega) = \sigma_0^2(\omega)$, white noise variances of the microphones $m_k(u)$ and $m_k(u+1)$ are

$$\sigma_U^2(\omega) = \sigma_0^2(\omega)/\left(2\left|h_c(\omega)\right|^2\right). \tag{25}$$

Output noise power is sum of contributions of omnidirectional and unidirectional microphones. Substituting Eq. (25) into Eq. (24) and adding $P_{n\_\mathrm{omni}}(\omega)$, the output noise power is

$$P_{n\_\mathrm{TOT}}(\omega)=\sigma_o^2\left[\sum_{m\epsilon O}|h_m(\omega)|^2+\frac{1}{2\,|h_c(\omega)|^2}\cdot\sum_{u\epsilon U}\left(\left|h_{k(u)}(\omega)\right|^2+\left|h_{k(u)+1}(\omega)\right|^2\right)\right], \quad (26)$$

where U is the set of unidirectional microphones of the microphone array. WNG is the ratio between input and output noise power expressed by (Bitzer, Simmer, 2001)

$$\mathrm{WNG}(\omega)=\frac{\sigma_o^2}{P_{n\_\mathrm{TOT}}(\omega)}=\frac{1}{a^*}, \quad (27)$$

where

$$a^*=\sum_{m\epsilon O}|h_m(\omega)|^2+\frac{1}{2\,|h_c(\omega)|^2}\sum_{m\epsilon U}\left(\left|h_{k(u)}\omega)\right|^2+\left|h_{k(u)+1}(\omega)\right|^2\right).$$

3. Results

In all tests we assume that the target signal is direct path speech for which the transfer vector $\mathbf{G}(\omega)$ is

$$\mathbf{G}(\omega)=\left[1\quad e^{-j\omega\tau_2}\quad\ldots\quad e^{-j\omega\tau_M}\right]^{\mathrm{T}},$$

where τ_m is time delay on m-th microphone, $m=2,\ldots,M$, relatively to the microphone m_1.

3.1. SNR gain in ideally diffuse noise

Due to the multipath effect, noise in a real room may have an energy flow of equal probability in all directions (Pan *et al.*, 2014; Bitzer, Simmer, 2001; McCowan, Bourlard, 2003). Hence, diffuse noise is usually used to predict the performance of the array in a real reverberant room. In diffuse noise, $(i,\,j)$-th element of the noise coherence matrix $\mathbf{\Gamma}_V(\omega)$ is (Pan *et al.*, 2014)

$$[\mathbf{\Gamma}_v(\omega)]_{i,j}=\sin[\omega(\tau_i-\tau_j)]/[\omega(\tau_i-\tau_j)]. \quad (28)$$

Substituting Eq. (28) into Eq. (22), we can calculate SNR gain in terms of frequency. SNR gains were evaluated on 4-microphone array for three analyzed configurations:

(i) an array composed of four omnidirectional microphones ("4-omni" configuration),

(ii) an array composed of one cardioid microphone and three omnidirectional microphones ("1-uni 3-omni" configuration),

(iii) an array composed of four cardioid microphones ("4-uni" configuration).

In all configurations, the distances between adjacent microphones were 5 cm. The distance between microphones in the differential microphone model was 1.5 cm. The selected distance of 1.5 cm is the half wavelength for 11 333 Hz. It provides a good accuracy of the unidirectional model for frequencies in the range from 100 to 8000 Hz. White noise gain was calculated by Eq. (27) and kept over 1/4, (i.e. −6 dB). Diagonal loading of the matrix $\mathbf{\Gamma}_v(\omega)$ is also applied by Eq. (8) with experimentally determined scalar δ. Figure 4 displays SNR gains for tested microphone array configurations calculated by Eq. (22).

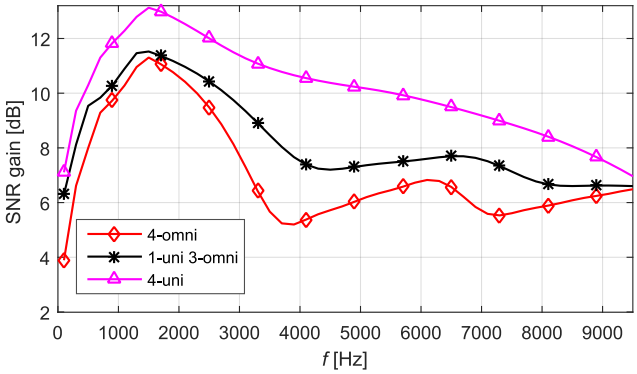

Fig. 4. Comparison of the SNR gains for different array configurations when WNG is limited to −6 dB.

3.2. Results in the simulated reverberant room

The performance of the different microphone array configurations was tested on the model of the small reverberant room shown in Fig. 5. The room was modeled by the acoustic image method (Allen, Berkley, 1979). Dimensions of the room were $5\times4\times2.85$ m. Reflection coefficients were 0.7754 for all walls. Reverberation time T_{60} was 400 ms calculated by Eyring's formula

$$T_{60}=0.163V/(-S\log\beta), \quad (29)$$

where V is the volume of the room [m^3], S is the total surface of the room [m^2], and β is reflection coefficient. Ambient noise was the sound of an air conditioner, https://www.soundsnap.com/tags/air_conditioner. To be more realistic, the air conditioner was represented by three independent point sources horizontally spaced 35 cm along the x-axis, as shown in Fig. 5. Speech signal was taken from the "Harvard Sentences" database, http://www.cs.columbia.edu/~hgs/audio/harvard.html. Processing was performed with a 16 kHz sampling frequency. We compared performances of the same array configurations analized in Subsec. 3.1.

The noise covariance matrix $\mathbf{\Phi}_v(\omega)$ was estimated in the pause of speech, while LCMV weights were cal-

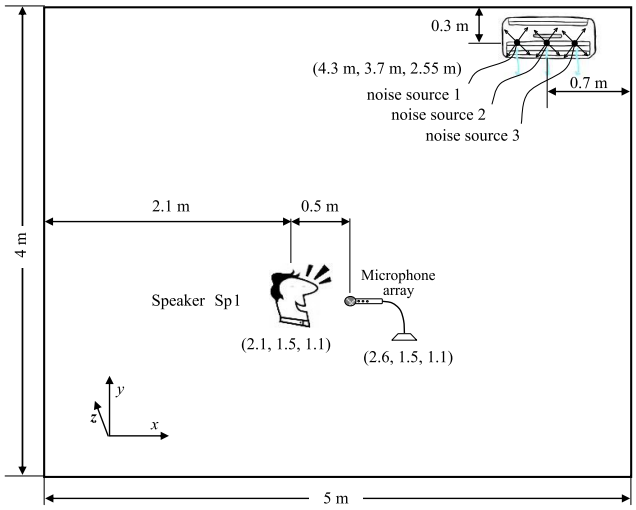

Fig. 5. Test scenario of the simulated room. Coordinates are x – width, y – depth, z – height.

culated by Eq. (7). Once estimated, these parameters were fixed and used to process microphone signals. The performance of the tested array configuration is evaluated by using three measures:

- **Noise attenuation** (NA) measure, which is defined by

$$A_{\text{noise}} = P_{v1}/P_{v\,\text{out}}, \tag{30}$$

where P_{v1} is noise power on reference microphone *mic* 1 calculated during the non-speech interval, $P_{v\,\text{out}}$ is noise power at the output of the beamformer calculated during the same non-speech interval. This measure displays the ability of the beamformer to suppress ambient noise under the assumption of unit gain for the desired speech.

- **Signal to noise ratio gain (SNR_gain) measure** takes into account the possible attenuation of the desired speech. This measure is defined by

$$\text{SNR\_gain} = o\text{SNR}/i\text{SNR}, \tag{31}$$

where iSNR is input signal to noise ratio,

$$i\text{SNR} = P_{s\_m1}/P_{v\_m1}. \tag{32}$$

P_{s_m1} is speech power on the reference microphone *mic* 1, P_{v_m1} is noise power at the same microphone, oSNR is output signal to noise ratio,

$$o\text{SNR} = P_{s\,\text{out}}/P_{v\,\text{out}}, \tag{33}$$

where $P_{s\,\text{out}}$ is speech power at the output of the beamformer, and $P_{v\,\text{out}}$ is noise at the output of the beamformer. SNR_gain is not sensitive to the suppression of the reverberation.

- **Signal to reverberation ratio** (SRR) is the ratio between the power of the direct path speech and the power of reflections at the output of the beamformer. SRR is calculated by

$$\text{SRR} = \frac{\sum_{t=t1}^{t2} \left(s_{dp\,\text{out}}(t)\right)^2}{\sum_{t=t1}^{t2} \left(s_{\text{out}}(t) - s_{dp\,\text{out}}(t)\right)^2}, \tag{34}$$

where $s_{dp\,\text{out}}(t)$ is direct path speech recorded at the output of the beamformer, and $s_{\text{out}}(t)$ is the total output of the beamformer in the reverberant room, (t_1, t_2) is speech interval.

In order to evaluate the robustness of the beamformer against small changes of the speaker's position, we compared its performances for three cases: (a) no steering error, (b) with steering error +15°, (c) with steering error −15°. WNG was limited to 1/4 (−6 dB). The results are displayed in Table 1. Quality measures are displayed in dB. We note that A_{noise} doesn't depend on the steering error because this measure doesn't take into account speech power attenuation at the output of the beamformer.

Table 1. Quality measures for WNG > 1/4 (−6 dB).

Measure	Algorithm	No steering error [dB]	Steering error +15° [dB]	Steering error −15° [dB]
A_{noise}	4-omnidirectional	10.472	10.472	10.472
	1-unidirectional/3-omnidirectional	13.193	13.193	13.193
	4-unidirectional	13.314	13.314	13.314
SNR_gain	4-omnidirectional	9.560	9.441	9.317
	1-unidirectional/3-omnidirectional	12.835	12.590	12.514
	4-unidirectional	12.483	11.995	11.951
SRR	4-omnidirectional	8.538	8.509	8.702
	1-unidirectional/3-omnidirectional	10.037	9.907	10.066
	4-unidirectional	11.436	11.045	11.192

3.3. Tests in the real room

Real room tests were aimed to evaluate the ability of the microphone array to suppress the ambient noise in the real room environment. Tests were conducted in a small office room. Dimensions of the room were $5 \times 4 \times 2.85$ m. The reverberation time was T_{60} = 400 ms exactly the same as in the scenario simulated in Subsec. 3.2. There were three active noises. The first was an air conditioner placed near the ceiling, which generated ambient noise of 45 dB. The second was the cooler of the notebook PC placed at a distance of 110 cm from the microphone array, which generated a noise level of 56 dB at the position of the microphone array. The third was the street noise of 40 dB SPL. The noise level was measured by sound level meter CESVA SC420. As in the simulation Subsec. 3.2, speech signals were taken from the "Harvard Sentences" database, http://www.cs.columbia.edu/~hgs/audio/harvard.html. The speech was played by loudspeaker placed at the distance of 90 cm in front of the microphone array generating 65 dB at the position of the microphone array, Fig. 6.

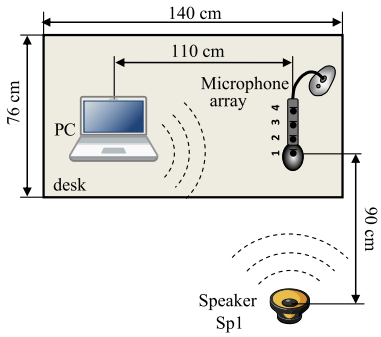

Fig. 6. Positions of the microphone array, PC and loudspeaker Sp1.

Acoustic signals were acquired by a conference microphone Proel BMG2 upgraded with 3 additional electret microphones (mic 2–mic 4), Fig. 7. Digitalization was performed with 48 kHz sampling rate. Data acquisition was conducted by software developed in MS Visual C++. The processing was performed in Matlab at 16 kHz sampling rate.

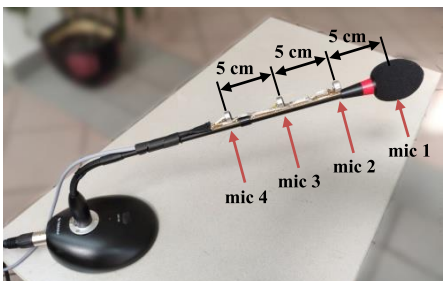

Fig. 7. Four-element microphone array composed of Proel conference microphone BMG2 (mic 1) and the additional 3 electret microphone capsules (mic 2, 3 and 4).

According to the (Ricketts, 2001), we have two options for the practical realization of the unidirectional microphones. Firstly, we can use two closely-spaced omnidirectional microphones and apply the model presented in Subsec. 2.3. Secondly, we can use a single unidirectional microphone capsule with the cardioid polar pattern. As the second option is cheaper and more practical for realization, we used it in real room experiments. Tests were conducted on two array configurations:

(a) "1-uni 3-omni" – an array composed of one unidirectional microphone and three omnidirectional electret microphones. The first microphone was the original BMG2 unidirectional microphone, while other microphones were WM 61A Panasonic.

(b) "4-uni" – an array composed of four unidirectional microphones. The first microphone was the same unidirectional microphone as in configuration (a), while the next three microphones were Coolvox MDN-318.

Test sentences were the same as in simulated room experiments. We generated 15 test examples which were independently processed by the proposed method. SNR_gain was obtained by comparing SNR of the tested array configuration with the SNR on the omnidirectional microphone (WM 61A).

SNR_gain of two tested array configurations and the SNR_gain on the single unidirectional microphone are displayed in Fig. 8. The average SNR_gain of the single unidirectional microphone was 4.39 dB with a standard deviation of 0.622. SNR_gains of the array configurations "1-uni 3-omni" and "4-uni" were 12.21 dB and 13.83 dB respectively. Corresponding standard deviations were 2.094 and 1.225 respectively.

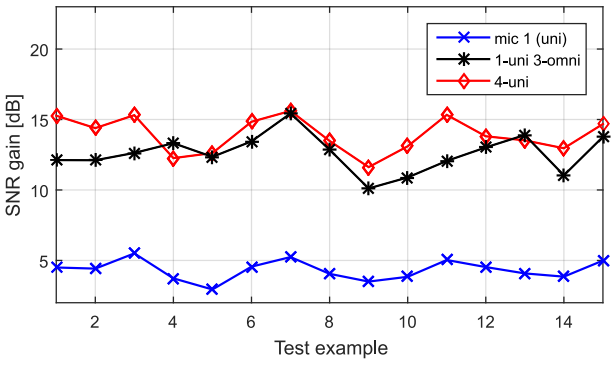

Fig. 8. SNR_gain of different configurations in the real room.

Speech quality was also evaluated by PESQ measure (ITU-T, 2001) using the corresponding speech samples from the database as reference. PESQ improvement was calculated relative to the PESQ at the output of the omnidirectional microphone. The results are displayed in Fig. 9. The average PESQ improve-

ments at the output of the single unidirectional microphone, at the output of the "1-uni 3-omni" array, and at output of the "4-uni" array were 0.41, 0.56, and 0.63, respectively. Corresponding standard deviations were 0.25, 0.24, and 0.23, respectively.

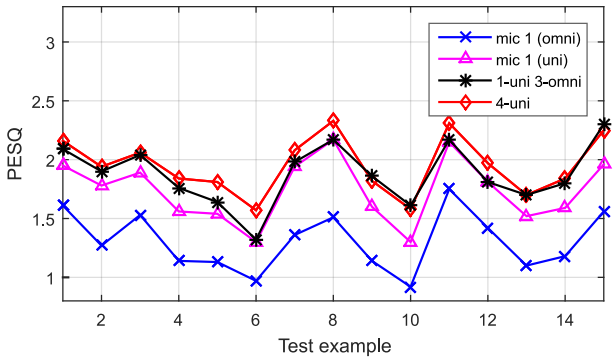

Fig. 9. PESQ measure of different configurations in the real room.

Typical time diagrams are displayed in Fig. 10 for the test sentence "The last switch cannot be turned off".

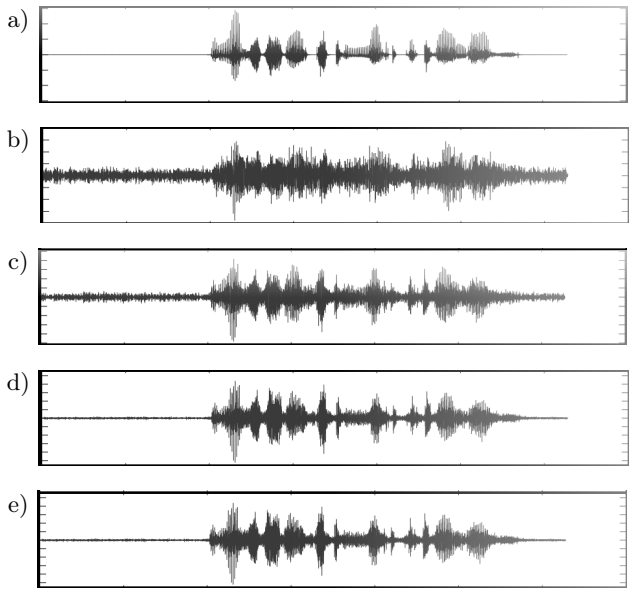

Fig. 10. Time diagrams: (a) original speech signal "The last switch cannot be turned off", (b) output of the single omnidirectional microphone, (c) output of the single unidirectional microphone, (d) output of the array configuration "1-omni 3-uni", (e) output of the array configuration "4-uni".

4. Discussion

The aim of the performance analysis of the proposed microphone array was to investigate its theoretical limits as well as to evaluate its performance in the real reverberant room. We were also interested in which combination of unidirectional and omnidirectional microphones maximally suppresses ambient noise and delivers the best perceptual speech quality.

Three types of noise sources are commonly used for microphone array analysis: (a) coherent noise, (b) diffuse noise, and (c) uncorrelated noise. Generally speaking, a linear array of N omnidirectional microphones can suppress coherent interferences coming from $N-1$ point sources (Van Trees, 2004). In addition to $N-1$ point sources, an array composed of N unidirectional microphones can suppress an additional interference coming from the direction of the spatial null of unidirectional microphone's beam pattern. For the cardioid microphone spatial null is on 180°.

Suppression of the diffuse noise depends on the array configuration. Maximal noise suppression, i.e., maximal SNR_gain, is obtained for the array composed of only unidirectional microphones (i.e., configuration denoted by "4-uni"), Fig. 4. The configuration "1-uni 3-omni" provides better SNR gain than the configuration "4-omni", but worse than the configuration "4-uni".

Evaluation of the SRR measure gives us useful information about the improvement of speech clarity. While the SNR_gain measure does not distinguish the power of the direct path from the power of the echo, the SRR measure explicitly estimates echo suppression. From the Table 1 we see that "4-uni" configuration has more than 1 dB better SRR than "1-uni 3-omni" configuration. Higher suppression of the reverberation by "4-uni" configuration explains its lower SNR_gain compared to the "1-uni 3-omni" array configuration.

Columns 2 and 3 in the Table 1 display performance of the proposed beamforming method in the presence of steering errors +15° and −15° respectively. Comparing columns 2 and 3 with column 1 (no steering error), we see that the SNR_gain of each array configuration is not degraded by more than 0.5 dB. Hence, we can say that the proposed microphone array with weights estimated in the pause of speech is robust against moderate steering errors (Saric, Jovicic, 2004).

Tests in real reverberant room evaluated the performance of the proposed microphone array model. Similar to the experiments with the simulated model of the reverberant room, the best SNR_gain is obtained by the "4-uni" configuration, Fig. 8. The obtained SNR_gain of 13.83 dB is the result of combined action of the cardioid microphone beam patterns and the directivity of the four-element LCMV microphone array. SNR gain of the "1-uni 3-omni" array configuration was 1.62 dB lower.

The proposed microphone array improves speech quality by two means. The first is by improving SNR, and the second is by dereverberation of the room. Perceptual speech quality was assessed by the PESQ measure. Results, displayed in Fig. 9, again showed the best performances of the "4-uni" array configuration.

5. Conclusions

In this paper, the end-fire adaptive microphone array, composed of an arbitrary number of unidirectional and omnidirectional microphones was analysed. Theoretical analysis was performed by modelling each unidirectional microphone with a two-element differential array and using an additional constraint of the LCMV beamformer for each unidirectional microphone. The analysis also included white noise gain estimation adapted for the analysed array configurations.

White noise gain limit, diagonal loading and adaptation in the pause of speech were included in the proposed beamforming method. Performance analysis of the various combinations of unidirectional and omnidirectional microphones conducted on the diffuse noise model, in the simulated and in the real reverberant room showed the best performance of the microphone array composed of only unidirectional microphones. The robustness analysis in terms of the steering errors showed a good performance of the proposed microphone array for moderate variations of the speaker's position.

The proposed processing method is linear with no additional distortion of the speech. Hence it can be successfully used in any application that demands high quality of the speech. It is worth noting that, around speech recognition threshold (SRT), even small gain in SNR leads to an appreciable increase in intelligibility, e.g., an SNR gain of 1 dB near SRT leads to an increase in the intelligibility of 5–10%, depending on the interference material (Wang, Brown, 2006, Fig. 1.3). Elder listeners and the subjects with hearing loss have higher SRT. For these people, even a single dB of the SNR gain can significantly improve their speech communication.

Acknowledgments

This paper is a result of research funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

1. Allen J.B., Berkley D.A. (1979), Image method for efficiently simulating small-room acoustics, *The Journal of the Acoustical Society of America*, **65**(4): 943–950, doi: 10.1121/1.382599.
2. Benesty J., Chen J., Huang Y. (2008), *Microphone Array Signal Processing*, Springer-Verlag, Berlin, doi: 10.1007/978-3-540-78612-2.
3. Bitzer J., Simmer K.U. (2001), Superdirective microphone arrays, [in:] *Microphone arrays. Digital Signal Processing*, Brandstein M., Ward D. [Ed.], pp. 19–38, Springer, Berlin, Heidelberg, doi: 10.1007/978-3-662-04619-7_2.
4. Capon J. (1969), High-resolution frequency-wavenumber spectrum analysis, *Proceedings of the IEEE*, **57**(8): 1408–1418, doi: 10.1109/PROC.1969.7278.
5. Chen J., Wang Y., Wang D. (2014), A feature study for classification-based speech separation at low signal-to-noise ratios, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **22**(12): 1993–2002, doi: 10.1109/ICASSP.2014.6854965.
6. DeFatta D.J. Lucas J.G., Hodgkiss W.S. (1988), *Digital Signal Processing: A System Design Approach*, John Wiley and Sons, Hoboken, New York, USA.
7. Elko G.W. (2004), Differential microphone arrays, [in:] *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Huang Y., Benesty J. [Eds], pp. 11–65, Springer, Boston, MA, USA, doi: 10.1007/1-4020-7769-6_2.
8. Frost O.L. (1972), An algorithm for linearly constrained adaptive array processing, *Proceedings of the IEEE*, **60**(8): 926–935, doi: 10.1109/PROC.1972.8817.
9. Greenberg J.E., Zurek P.M. (2001), Microphone-array hearing aids, [in:] *Microphone Arrays*, Brandstein M., Ward D. [Eds], pp. 229–253, Springer, Berlin, Heidelberg, doi: 10.1007/978-3-662-04619-7_11.
10. Hoshuyama O., Sugiyama A., Hirano A. (1999), A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters, *IEEE Transactions on signal processing*, **47**(10): 2677–2684, doi: 10.1109/78.790650.
11. International Telecommunications Union [ITU-T] (2001), *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, Recommendation P.862 (02/01).
12. Jovicic S.T., Saric Z.M., Turajlic S.R. (2005), Application of the maximum signal to interference ratio criterion to the adaptive microphone array, *Acoustics Research Letters Online*, **6**(4): 232–237, doi: 10.1121/1.1989785.
13. Kates J.M., Weiss M.R. (1996), A comparison of hearing-aid array-processing techniques, *The Journal of the Acoustical Society of America*, **99**(5): 3138–3148, doi: 10.1121/1.414798.
14. Krecichwost M. Miodonska Z., Badura P., Trzaskalik J., Mocko N. (2019), Multi-channel acoustic analysis of phoneme /s/ mispronunciation for lateral sigmatism detection, *Biocybernetics and Biomedical Engineering*, **39**(1): 246–255, doi: 10.1016/j.bbe.2018.11.005.
15. Krecichwost M., Miodonska Z., Trzaskalik J., Badura P. (2020), Multichannel speech acquisition and analysis for computer-aided sigmatism diagnosis in children, *IEEE Access*, **8**: 98647–98658, doi: 10.1109/ACCESS.2020.2996413.
16. Marro C., Mahieux Y., Simmer K.U. (1998), Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering, *IEEE Transactions on Speech and Audio Processing*, **6**(3): 240–259, doi: 10.1109/ACCESS.2020.2996413.

17. McCowan I.A., Bourlard H. (2003), Microphone array post-filter based on noise field coherence, *IEEE Transactions on Speech and Audio Processing*, **11**(6): 709–716, doi: 10.1109/TSA.2003.818212.

18. McDonough J., Kumatani K. (2012), Microphone arrays, [in:] *Techniques for Noise Robustness in Automatic Speech Recognition*, Virtanen T. [Ed.], pp. 109–157, John Wiley and Sons, Hoboken, NJ, USA, doi: 10.1002/9781118392683.ch6.

19. Pan C., Chen J., Benesty J. (2014), On the noise reduction performance of the MVDR beamformer in noisy and reverberant environments, *Proceedings 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 815–819, Florence, doi: 10.1109/ICASSP.2014.6853710.

20. Pan C., Chen J., Benesty J. (2015), A multistage minimum variance distortionless response beamformer for noise reduction, *The Journal of the Acoustical Society of America*, **137**(3): 1377–1388, doi: 10.1121/1.4913459.

21. Papp I.I., Saric Z.M., Jovicic S.T., Teslic N.D. (2007), Adaptive microphone array for unknown desired speaker's transfer function, *The Journal of the Acoustical Society of America*, **122**(2): EL44–EL49, doi: 10.1121/1.2749077.

22. Papp I.I., Saric Z.M., Teslic N.D. (2011), Handsfree voice communication with TV, *IEEE Transactions on Consumer Electronics*, **57**(2): 606–614, doi: 10.1109/TCE.2011.5955198.

23. Parra L., Spence C. (2000), Convolutive blind separation of non-stationary sources, *IEEE transactions on Speech and Audio Processing*, **8**(3): 320–327, doi: 10.1109/89.841214.

24. Parra L.C., Alvino C.V. (2002), Geometric source separation: Merging convolutive source separation with geometric beamforming, *IEEE Transactions on Speech and Audio Processing*, **10**(6): 352–362, doi: 10.1109/TSA.2002.803443.

25. Ricketts T.A. (2001). Directional hearing aids, *Trends in Amplification*, **5**(4): 139–176, doi: 10.1177/108471380100500401.

26. Saric Z.M., Jovicic S.T. (2004), Adaptive microphone array based on pause detection, *Acoustics Research Letters Online*, **5**(2): 68–74, doi: 10.1121/1.1650411.

27. Saric Z.M., Simic D.P., Jovicic S.T. (2011), A new post-filter algorithm combined with two-step adaptive beamformer, *Circuits, Systems, and Signal Processing*, **30**(3): 483–500, doi: 10.1007/s00034-010-9233-1.

28. Šarić Z., Subotić M., Bilibajkić R., Barjaktarović M. (2019), Bidirectional microphone array with adaptation controlled by voice activity detector based on multiple beamformers, *Multimedia Tools and Applications*, **78**(11): 15235–15254, doi: 10.1007/s11042-018-6895-3.

29. Simmer K.U., Bitzer J., Marro C. (2001), Post-filtering techniques, [in:] *Microphone Arrays. Digital Signal Processing*, Brandstein M., Ward D. [Eds], pp. 39–60, Springer, Berlin, Heidelberg, doi: 10.1007/978-3-662-04619-7_3.

30. Soede W., Berkhout A.J., Bilsen F.A. (1993), Development of a directional hearing instrument based on array technology, *The Journal of the Acoustical Society of America*, **94**(2): 785–798, doi: 10.1121/1.408180.

31. Spriet A., Moonen M., Wouters J. (2002), A multi-channel subband generalized singular value decomposition approach to speech enhancement, *European Transactions on Telecommunications*, **13**(2): 149–158, doi: 10.1002/ett.4460130210.

32. Trucco A., Traverso F., Crocco M. (2015), Maximum constrained directivity of oversteered end-fire sensor arrays, *Sensors*, **15**(6): 13477–13502, doi: 10.3390/s150613477.

33. Van Trees H.L. (2004), *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, John Wiley and Sons, Hoboken, NJ, USA, doi: 10.1002/0471221104.

34. Wang D.L., Brown G.J.6 (2006), Fundamentals of computational auditory scene analysis, [in:] *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*, Wang D.L., Brown G.J. [Eds], John Wiley and Sons, Hoboken, NJ, pp. 1–37, doi: 10.1109/9780470043387.ch1.

35. Wang L., Ding H., Yin F. (2010), Combining superdirective beamforming and frequency-domain blind source separation for highly reverberant signals, *EURASIP Journal on Audio, Speech, and Music Processing*, **4**: 1–13, doi: 10.1155/2010/797962.

36. Wang D., Chen J. (2018), Supervised speech separation based on deep learning: An overview, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **26**(10): 1702–1726, doi: 10.1109/TASLP.2018.2842159.

37. Wang X., Cohen I., Chen J., Benesty J. (2019), On robust and high directive beamforming with small-spacing microphone arrays for scattered sources, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, **27**(4): 842–852, doi: 10.1109/TASLP.2019.2899517.

38. Wölfel M., McDonough J.W. (2009), *Distant Speech Recognition*, John Wiley and Sons, Hoboken, NJ, doi: 10.1002/9780470714089.

39. Yilmaz O., Rickard S. (2004), Blind separation of speech mixtures via time-frequency masking, *IEEE Transactions on signal processing*, **52**(7): 1830–1847, doi: 10.1109/TSP.2004.828896.

40. Zelinski R. (1988), A microphone array with adaptive post-filtering for noise reduction in reverberant rooms, *Proceedings of ICASSP-88 International Conference on Acoustics, Speech, and Signal Processing*, **5**: 2578–2579, doi: 10.1109/ICASSP.1988.197172.