

PERFORMANCE ANALYSIS OF THE EXTENDED BINAURAL MVDR BEAMFORMER WITH PARTIAL NOISE ESTIMATION IN A HOMOGENEOUS NOISE FIELD

Nico Gößling, Daniel Marquardt, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics and
Cluster of Excellence Hearing4All, Oldenburg, Germany
{nico.goessling, daniel.marquardt, simon.doclo}@uni-oldenburg.de

ABSTRACT

Besides noise reduction an important objective of a binaural speech enhancement algorithm is the preservation of the binaural cues of both the desired speech source as well as the undesired noise sources. Recently, it has been shown that by combining a binaural hearing system with an external microphone it is possible to improve both the noise reduction performance as well as the binaural cue preservation for a directional noise source. In this paper, we consider the binaural minimum variance distortionless response beamformer with partial noise estimation (MVDR-N), which aims at preserving the speech component and a scaled version of the noise component. For a homogeneous noise field, we theoretically analyze the influence of an external microphone on the performance of the binaural MVDR-N beamformer, showing that a larger SNR improvement can be achieved and a smaller scaling parameter is required to achieve a desired interaural coherence for the residual noise. Experimental simulation results in a realistic acoustic scenario confirm these theoretical findings.

Index Terms— Acoustic sensor networks, noise reduction, binaural cues, interaural coherence

1. INTRODUCTION

Noise reduction algorithms for head-mounted assistive listening devices (e.g., hearing aids) are crucial to improve speech quality and intelligibility in background noise. Binaural devices, consisting of one or more microphones on each side of the head of the listener, are able to exploit not only spectral but also spatial information [1–3]. Besides reducing noise, preserving the binaural cues of all sound sources is an important objective of a binaural noise reduction algorithm in order to make sure that the listener’s impression of the acoustic scene is not distorted.

In case of a single desired speech source, it has been shown that the binaural multi-channel Wiener filter (MWF) [2, 4] preserves the binaural cues of the speech source, but typically distorts the binaural cues of the noise, such that the residual noise is perceived as coming from the same direction as the speech source. Hence, the binaural MWF with partial noise estimation (MWF-N) has been proposed [2, 4, 5], which aims at preserving the speech component and a scaled version of the noise component in the reference microphones of the left and the right hearing device. It has been shown that the scaling parameter in the binaural MWF-N allows to trade off noise reduction performance and binaural cue preservation of the noise component [4]. For a homogeneous noise field (e.g., diffuse noise) a closed-form expression for the scaling parameter yielding

a desired interaural coherence (IC) of the residual noise has been derived in [6].

In general, acoustic sensor networks consisting of spatially distributed microphones enable to exploit a larger spatial diversity than traditional microphone arrays [7, 8]. Hence, the advantages of using one or more external microphones in combination with (head-mounted) communication devices have recently been explored [9–12]. For a directional noise source, it has been shown for the binaural MWF-N that incorporating an external microphone into a binaural hearing system enables to improve both the noise reduction performance as well as the binaural cues (i.e., interaural time and level difference) of the residual noise [12].

In this paper, we analyze the influence of an external microphone on the performance of the binaural minimum variance distortionless response beamformer with partial noise estimation (MVDR-N), which is a special case of the binaural MWF-N [13], for a homogeneous noise field. We first derive general theoretical expressions, showing that (using an external microphone) a larger SNR improvement can be achieved and a smaller scaling parameter is required to achieve a desired IC for the residual noise. Experimental simulation results using a psycho-acoustically motivated desired IC show that a larger SNR improvement can be achieved depending on the distance between the desired speech source and the external microphone, even in the presence of estimation errors.

2. CONFIGURATION AND NOTATION

Consider the binaural noise reduction system in Fig. 1, consisting of a head-mounted microphone array with M_L microphones on the left side, a head-mounted microphone array with M_R microphones on the right side, and an external microphone. Considering an acoustic scenario with a single desired speech source in a homogeneous noise field, the m -th microphone signal of the left and the right microphone array in the frequency-domain is given by

$$Y_{\{L,R\},m}(\omega) = X_{\{L,R\},m}(\omega) + N_{\{L,R\},m}(\omega), \quad (1)$$

with $X_{\{L,R\},m}(\omega)$ denoting the speech component and $N_{\{L,R\},m}(\omega)$ the noise component. For conciseness, the frequency variable ω will be omitted in the remainder of the paper. All microphone signals of the left and the right microphone array can be stacked in an M -dimensional vector, with $M = M_L + M_R$, as

$$\mathbf{y} = [Y_{L,1} \dots Y_{L,M_L} \ Y_{R,1} \dots Y_{R,M_R}]^T, \quad (2)$$

which can be written as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad (3)$$

where the vectors \mathbf{x} and \mathbf{n} are defined similarly as in (2). For a single speech source, the speech vector is given by

$$\mathbf{x} = S_x \mathbf{a}, \quad (4)$$

This work was supported in part by the Cluster of Excellence 1077 “Hearing4All”, funded by the German Research Foundation (DFG), and by the joint Lower Saxony-Israeli Project ATHENA.

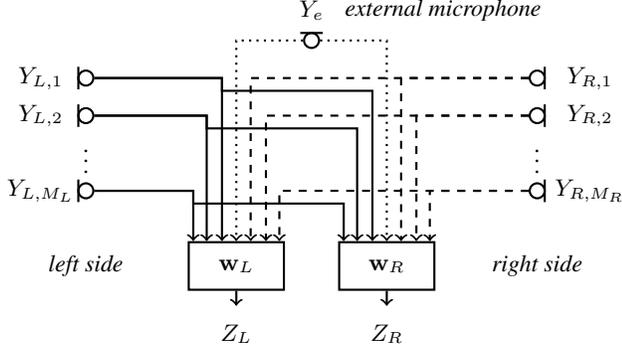


Fig. 1: Binaural noise reduction system in combination with an external microphone.

with S_x the speech signal and \mathbf{a} the acoustic transfer function (ATF) vector between the microphones of the head-mounted arrays and the speech source. Without loss of generality, we define the first microphone of both arrays as the reference microphones. The reference microphone signals $Y_{L,1}$ and $Y_{R,1}$ are denoted as Y_L and Y_R , i.e.,

$$Y_L = \mathbf{e}_L^T \mathbf{y}, \quad Y_R = \mathbf{e}_R^T \mathbf{y}, \quad (5)$$

where \mathbf{e}_L and \mathbf{e}_R denote M -dimensional selection vectors with all elements equal to 0 except one element equal to 1. Hence, the reference microphone signals can be written as

$$Y_L = S_x A_L + N_L, \quad Y_R = S_x A_R + N_R. \quad (6)$$

The relative transfer function (RTF) vectors for the left and the right array are defined as

$$\bar{\mathbf{a}}_L = \frac{\mathbf{a}}{A_L}, \quad \bar{\mathbf{a}}_R = \frac{\mathbf{a}}{A_R}. \quad (7)$$

The $M \times M$ -dimensional correlation matrices of the speech and the noise components are defined as

$$\mathbf{R}_x = \mathcal{E} \{ \mathbf{x} \mathbf{x}^H \} = \Phi_s \mathbf{a} \mathbf{a}^H, \quad \mathbf{R}_n = \mathcal{E} \{ \mathbf{n} \mathbf{n}^H \}, \quad (8)$$

where $\mathcal{E} \{ \cdot \}$ denotes the expectation operator, H the conjugate transpose and $\Phi_s = \mathcal{E} \{ |S_x|^2 \}$ the power spectral density (PSD) of the speech signal. The noise correlation matrix is assumed to be full rank and hence invertible. Assuming statistical independence between \mathbf{x} and \mathbf{n} , the correlation matrix of the microphone signals $\mathbf{R}_y = \mathcal{E} \{ \mathbf{y} \mathbf{y}^H \}$ can be written as

$$\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_n. \quad (9)$$

The PSD and the cross-power spectral density (CPSD) of the speech component in the reference microphone signals are given by

$$\Phi_{x,\{L,R\}} = \mathcal{E} \{ |X_{\{L,R\}}|^2 \} = \Phi_s |A_{\{L,R\}}|^2, \quad (10)$$

$$\Phi_{x,LR} = \mathcal{E} \{ X_L X_R^* \} = \Phi_s A_L A_R^*. \quad (11)$$

Using the assumption of a homogeneous noise field, the PSD of the noise component in the reference microphone signals is given by

$$\Phi_n = \mathcal{E} \{ |N_L|^2 \} = \mathcal{E} \{ |N_R|^2 \}. \quad (12)$$

All elements of \mathbf{R}_n are assumed to be real-valued. The (binaural) output signals of the left and the right microphone array Z_L and Z_R , are obtained by summing filtered versions of all microphone signals (cf. Fig. 1), i.e.,

$$Z_{\{L,R\}} = \mathbf{w}_{\{L,R\}}^H \mathbf{y} = \mathbf{w}_{\{L,R\}}^H \mathbf{x} + \mathbf{w}_{\{L,R\}}^H \mathbf{n}, \quad (13)$$

with M -dimensional (complex-valued) filters \mathbf{w}_L and \mathbf{w}_R .

The input signal-to-noise ratio (SNR) of the left and the right

reference microphone signal is given by

$$SNR_L^{\text{in}} = \frac{\Phi_{x,L}}{\Phi_n}, \quad SNR_R^{\text{in}} = \frac{\Phi_{x,R}}{\Phi_n}. \quad (14)$$

The output SNR of the left and the right output signal is given by the ratio of the output PSD of the speech component and the output PSD of the noise component, i.e.,

$$SNR_L^{\text{out}} = \frac{\mathbf{w}_L^H \mathbf{R}_x \mathbf{w}_L}{\mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_L}, \quad SNR_R^{\text{out}} = \frac{\mathbf{w}_R^H \mathbf{R}_x \mathbf{w}_R}{\mathbf{w}_R^H \mathbf{R}_n \mathbf{w}_R}. \quad (15)$$

The (real-valued) input IC of the noise component is defined as the normalized cross-correlation between the reference microphones, i.e.,

$$IC_n^{\text{in}} = \frac{\mathcal{E} \{ N_L N_R^* \}}{\sqrt{\mathcal{E} \{ |N_L|^2 \} \mathcal{E} \{ |N_R|^2 \}}} = \frac{\mathbf{e}_L^T \mathbf{R}_n \mathbf{e}_R}{\Phi_n}. \quad (16)$$

The output IC of the noise component is defined as the normalized cross-correlation between the output components, i.e.,

$$IC_n^{\text{out}} = \frac{\mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_R}{\sqrt{(\mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_L)(\mathbf{w}_R^H \mathbf{R}_n \mathbf{w}_R)}}. \quad (17)$$

3. BINAURAL NOISE REDUCTION ALGORITHMS

In this section we briefly review two well-known binaural noise reduction algorithms: the binaural MVDR beamformer [2, 14] and the binaural MVDR-N beamformer [2, 4, 5, 13].

3.1. Binaural MVDR beamformer

The binaural MVDR beamformer aims at minimizing the output noise PSD while preserving the speech component in the left and the right reference microphone signals. Hence, the optimization problem for both filters is given by [2, 14, 15]

$$\min_{\mathbf{w}_{\{L,R\}}} \mathcal{E} \left\{ \left| \mathbf{w}_{\{L,R\}}^H \mathbf{n} \right|^2 \right\} \quad \text{s.t.} \quad \mathbf{w}_{\{L,R\}}^H \mathbf{x} = X_{\{L,R\}}. \quad (18)$$

Using (4), (7) and (8), the resulting filters can be written in terms of the RTF vectors as

$$\mathbf{w}_{\text{MVDR},\{L,R\}} = \frac{\mathbf{R}_n^{-1} \bar{\mathbf{a}}_{\{L,R\}}}{\bar{\mathbf{a}}_{\{L,R\}}^H \mathbf{R}_n^{-1} \bar{\mathbf{a}}_{\{L,R\}}}. \quad (19)$$

By substituting (19) into (15), the output SNR of the binaural MVDR beamformer for both the left and the right microphone array is equal to [2, 4]

$$SNR_{\text{MVDR}}^{\text{out}} = \rho_0 = \Phi_s \bar{\mathbf{a}}^H \mathbf{R}_n^{-1} \bar{\mathbf{a}}, \quad (20)$$

which is always larger than or equal to the input SNR in (14). As shown in [4, 16], the binaural MVDR beamformer preserves the binaural cues of the speech source, but distorts the IC of the noise in such a way that the speech and the residual noise are perceived as coming from the same direction, which is undesired.

3.2. Binaural MVDR beamformer with partial noise estimation

In order to better preserve the IC of the noise and hence the impression of the acoustic scene, the binaural MVDR-N beamformer has been proposed [2, 4, 5, 13], which aims at preserving the speech component and a scaled version of the noise component in the reference microphone signals. The optimization problem for both filters is given by

$$\min_{\mathbf{w}_{\{L,R\}}} \mathcal{E} \left\{ \left| \mathbf{w}_{\{L,R\}}^H \mathbf{n} - \eta N_{\{L,R\}} \right|^2 \right\} \quad \text{s.t.} \quad \mathbf{w}_{\{L,R\}}^H \mathbf{x} = X_{\{L,R\}}, \quad (21)$$

where η denotes a (real-valued) scaling parameter, with $0 \leq \eta \leq 1$. It has been shown that the resulting filters can be written as [13]

$$\mathbf{w}_{\text{MVDR-N},\{L,R\}}(\eta) = (1 - \eta)\mathbf{w}_{\text{MVDR},\{L,R\}} + \eta\mathbf{e}_{\{L,R\}}, \quad (22)$$

i.e., the output signals are equal to the sum of the binaural MVDR output signals (scaled with $1 - \eta$) and the reference microphone signals (scaled with η). Hence, for $\eta = 1$ the noise IC is completely preserved, whereas for $\eta = 0$ the noise IC is distorted in the same way as for the MVDR beamformer.

The output SNR of the binaural MVDR-N beamformer is given by [4, 13]

$$SNR_{\text{MVDR-N},\{L,R\}}^{\text{out}}(\eta) = \frac{\rho_0}{1 + \eta^2 \left(\frac{\rho_0}{SNR_{\{L,R\}}^{\text{in}}} - 1 \right)}, \quad (23)$$

where $\frac{\rho_0}{SNR_{\{L,R\}}^{\text{in}}} \geq 1$ represents the SNR improvement of the binaural MVDR beamformer. It can be easily seen that (23) is monotonically decreasing with η [4]. Hence, a larger η leads to a smaller output SNR of the binaural MVDR-N beamformer, thereby trading off binaural cue preservation of the noise component against noise reduction performance.

4. INFLUENCE OF EXTERNAL MICROPHONE

In this section the influence of incorporating an external microphone (cf. Fig. 1) on the output SNR and the output noise IC is investigated. Adding the external microphone signal $Y_e = X_e + N_e$ to the stacked input vector in (2) yields the $(M + 1)$ -dimensional extended input vector

$$\mathbf{y}_e = \begin{bmatrix} \mathbf{y} \\ Y_e \end{bmatrix}. \quad (24)$$

Similarly to (4), the speech component in the external microphone signal is equal to $X_e = S_x A_e$, with A_e the ATF between the speech source and the external microphone. The extended speech and noise component vectors \mathbf{x}_e and \mathbf{n}_e are defined similarly to (24). Hence, the extended speech and noise correlation matrices are given by

$$\mathbf{R}_{\mathbf{x},e} = \mathcal{E} \left\{ \mathbf{x}_e \mathbf{x}_e^H \right\} = \Phi_s \mathbf{a}_e \mathbf{a}_e^H, \quad (25)$$

$$\mathbf{R}_{\mathbf{n},e} = \mathcal{E} \left\{ \mathbf{n}_e \mathbf{n}_e^H \right\} = \begin{bmatrix} \mathbf{R}_n & \mathbf{r}_{n,e} \\ \mathbf{r}_{n,e}^H & \Phi_{n,e} \end{bmatrix}, \quad (26)$$

with $\mathbf{r}_{n,e} = \mathcal{E} \{ \mathbf{n} N_e^* \}$ the M -dimensional cross-correlation vector between the noise components in the microphone array signals and the external microphone signal and $\Phi_{n,e} = \mathcal{E} \{ |N_e|^2 \}$ the noise PSD in the external microphone signal.

Similarly to (7), the extended RTF vectors are defined as

$$\bar{\mathbf{a}}_{L,e} = \frac{\mathbf{a}_e}{A_L}, \quad \bar{\mathbf{a}}_{R,e} = \frac{\mathbf{a}_e}{A_R}. \quad (27)$$

Similarly to (19) and (22), the extended MVDR-N (eMVDR-N) beamformer is now given by

$$\mathbf{w}_{\text{eMVDR-N},\{L,R\}}(\eta) = (1 - \eta) \underbrace{\frac{\mathbf{R}_{n,e}^{-1} \bar{\mathbf{a}}_{\{L,R\},e}}{\bar{\mathbf{a}}_{\{L,R\},e}^H \mathbf{R}_{n,e}^{-1} \bar{\mathbf{a}}_{\{L,R\},e}}}_{\mathbf{w}_{\text{eMVDR},\{L,R\}}} + \eta \mathbf{e}_{\{L,R\},e}, \quad (28)$$

where the inverse of (26) can be written in terms of \mathbf{R}_n^{-1} as [17]

$$\mathbf{R}_{n,e}^{-1} = \begin{bmatrix} \mathbf{R}_n^{-1} + \xi \mathbf{R}_n^{-1} \mathbf{r}_{n,e} \mathbf{r}_{n,e}^H \mathbf{R}_n^{-1} & -\xi \mathbf{R}_n^{-1} \mathbf{r}_{n,e} \\ -\xi \mathbf{r}_{n,e}^H \mathbf{R}_n^{-1} & \xi \end{bmatrix}, \quad (29)$$

with

$$\xi = (\Phi_{n,e} - \mathbf{r}_{n,e}^H \mathbf{R}_n^{-1} \mathbf{r}_{n,e})^{-1} \quad (30)$$

the inverse Schur complement of \mathbf{R}_n in (26).

4.1. Influence of external microphone on output SNR

Similarly to (20), the output SNR of the eMVDR beamformer (i.e., $\eta = 0$) is given by

$$SNR_{\text{eMVDR},\{L,R\}}^{\text{out}} = \rho_e = \Phi_s \mathbf{a}_e^H \mathbf{R}_{n,e}^{-1} \mathbf{a}_e. \quad (31)$$

By substituting (29) into (31), the output SNR of the eMVDR beamformer is equal to

$$\rho_e = \rho_0 + \xi \Phi_s \left| \mathbf{r}_{n,e}^H \mathbf{R}_n^{-1} \mathbf{a}_e - A_e \right|^2 \quad (32)$$

Since $\mathbf{R}_{n,e}$ is positive definite, ξ in (30) is always positive [17], such that as expected the output SNR of the eMVDR beamformer ρ_e with an external microphone is always larger than or equal to the output SNR of the MVDR beamformer ρ_0 , i.e.,

$$\rho_e \geq \rho_0. \quad (33)$$

Similarly to (23) the output SNR of the eMVDR-N beamformer is equal to

$$SNR_{\text{eMVDR-N},\{L,R\}}^{\text{out}}(\eta) = \frac{\rho_e}{1 + \eta^2 \left(\frac{\rho_e}{SNR_{\{L,R\}}^{\text{in}}} - 1 \right)}. \quad (34)$$

Assuming a fixed acoustic scenario and head-mounted microphone configuration where only the position of the external microphone changes, ρ_0 can be considered as a constant and ρ_e as a variable solely depending on the external microphone position. By calculating the partial derivative with respect to ρ_e , it can be shown that the output SNR of the eMVDR-N beamformer is monotonically increasing with ρ_e , i.e.,

$$\frac{\partial SNR_{\text{eMVDR-N},\{L,R\}}^{\text{out}}(\eta)}{\partial \rho_e} = \frac{1 - \eta^2}{\left(1 - \eta^2 + \eta^2 \frac{\rho_e}{SNR_{\{L,R\}}^{\text{in}}} \right)^2} \geq 0. \quad (35)$$

Therefore, since $\rho_e \geq \rho_0$, the output SNR of the eMVDR-N beamformer is always larger than or equal to the output SNR of the MVDR-N beamformer for a fixed η , i.e.,

$$SNR_{\text{eMVDR-N},\{L,R\}}^{\text{out}}(\eta) \geq SNR_{\text{MVDR-N},\{L,R\}}^{\text{out}}(\eta) \quad (36)$$

Hence, by using an external microphone, either a larger output SNR can be obtained for the same scaling parameter η , or the same output SNR can be obtained for a larger scaling parameter, which is in line with the findings in [12], where a similar statement was made for a directional noise source.

4.2. Influence of external microphone on output noise IC

Assuming the desired speech source to be in front of the listener, i.e. $A_L = A_R$, such that $\Phi_x = \Phi_{x,L} = \Phi_{x,R}$ and $SNR^{\text{in}} = SNR_L^{\text{in}} = SNR_R^{\text{in}}$, the output noise IC of the eMVDR-N beamformer is given by [6]

$$IC_{n,\text{eMVDR-N}}^{\text{out}} = \frac{\frac{1 - \eta^2}{\rho_e} \Phi_x + \eta^2 \Phi_n IC_n^{\text{in}}}{\frac{1 - \eta^2}{\rho_e} \Phi_x + \eta^2 \Phi_n}. \quad (37)$$

It has been shown in [6] that the value of the scaling parameter yielding a desired (real-valued) noise interaural coherence IC_n^{des} , with $IC_n^{\text{in}} \leq IC_n^{\text{des}} \leq 1$, is equal to

$$\eta_{\text{des}}(\rho_e) = \frac{1}{\sqrt{\frac{IC_n^{\text{des}} - IC_n^{\text{in}}}{1 - IC_n^{\text{des}}} \frac{\rho_e}{SNR^{\text{in}}} + 1}}. \quad (38)$$

By calculating the partial derivative of (38) with respect to ρ_e , it can be shown that η_{des} is monotonically decreasing with ρ_e , i.e.,

$$\frac{\partial \eta_{\text{des}}(\rho_e)}{\partial \rho_e} = -\frac{\beta}{2(\beta \rho_e + 1)^{3/2}} \leq 0, \quad (39)$$

with

$$\beta = \frac{IC_n^{\text{des}} - IC_n^{\text{in}}}{1 - IC_n^{\text{des}}} \frac{1}{SNR^{\text{in}}} \geq 0. \quad (40)$$

Hence, by using an external microphone (for which $\rho_e \geq \rho_0$), the same desired output noise IC can be achieved with a smaller η , i.e.,

$$\eta_{\text{des}}(\rho_e) \leq \eta_{\text{des}}(\rho_0) \quad (41)$$

Therefore, for a homogeneous noise field the eMVDR-N beamformer provides a better trade-off between binaural cue preservation of the noise component and noise reduction performance than the MVDR-N beamformer, which is also in line with the findings in [12] for a directional noise source.

5. EXPERIMENTAL RESULTS

In this section, we present experimental results investigating the influence of an external microphone on the performance, i.e. SNR improvement and optimal scaling parameter, of the eMVDR-N beamformer. For the assumed acoustic scenario, i.e. a single speech source in front of the listener in a homogeneous noise field, we present theoretical validation results of the derived expressions in Section 4 as well as realistic simulation results.

For the theoretical validation we assume $SNR^{\text{in}} = 1$ and the output SNR of the binaural MVDR beamformer (without external microphone) equal to $\rho_0 = 5$. Fig. 2 (left) depicts the SNR improvement $\Delta SNR = SNR^{\text{out}}/SNR^{\text{in}}$ of the eMVDR-N beamformer, corresponding to (23), as a function of ρ_e/ρ_0 . Please note that $\rho_e/\rho_0 = 1$ corresponds to the performance of the MVDR-N beamformer (without external microphone). It can be observed that for all η the SNR improvement is monotonically increasing with ρ_e , cf. derivations in 4.1. For $IC_n^{\text{des}} = 0.6$, Fig. 2 (right) depicts the scaling parameter η_{des} leading to IC_n^{des} , corresponding to (38), for different IC_n^{in} . It can be observed that for all IC_n^{in} the optimal scaling parameter η_{des} is monotonically decreasing with ρ_e , cf. derivations in 4.2.

As a more realistic simulation we considered a room with dimensions $6 \text{ m} \times 6 \text{ m} \times 3 \text{ m}$ and reverberation time 300 ms, where the room impulse responses between the speech source and all microphones were simulated using the image method [18], modelling the head as a rigid sphere [19] with a radius of 8.5 cm. The listener and the speech source were placed at a distance of 3 m to each other and were surrounded by a spherically isotropic noise field. We considered $M_L = M_R = 2$ microphones on each side of the head with an inter-microphone distance of 7 mm. The external microphone was placed between the listener and the speech source at a distance from the speech source ranging between 0.25 m and 2.75 m. As a speech source signal a recording of a male speaker was used. The homogeneous noise field was generated using a multi-talker signal and using the method presented in [20]. The broadband input SNR at the reference microphones was equal to 0 dB. All simulations were performed at a sampling frequency of 16 kHz using an STFT-based weighted overlap-add framework with a frame length of 32 ms, frame shift of 50% and a Hanning window. The input signals consisted of 2 s noise-only, followed by 11 s speech-plus-noise. To take estimation errors into account, the noise correlation matrix $\hat{\mathbf{R}}_n$ was estimated from the noise-only part, while the speech correlation matrix $\hat{\mathbf{R}}_x$ was estimated from the speech-plus-noise part as $\hat{\mathbf{R}}_x = \hat{\mathbf{R}}_y - \hat{\mathbf{R}}_n$ with additionally using a rank-1 approximation [21]. The RTF vectors in (27) were estimated using a generalized eigenvalue decomposition-based method [22]. For the desired interaural coherence of the noise component IC_n^{des} , we have used the frequency-dependent values proposed in [16], which are psycho-

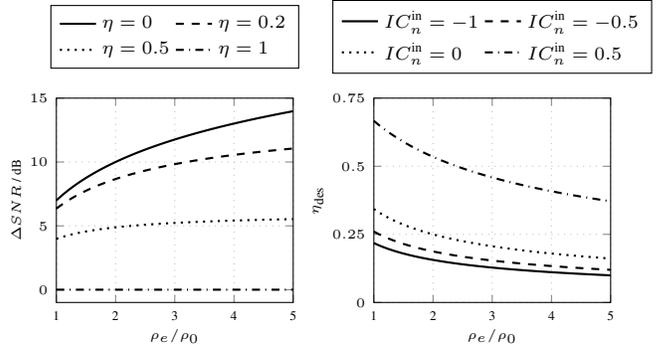


Fig. 2: Performance of the eMVDR-N beamformer: SNR improvement for different η (left) and optimal scaling parameter η_{des} for different IC_n^{in} (right), as a function of ρ_e/ρ_0 ($SNR^{\text{in}} = 1$, $\rho_0 = 5$, $IC_n^{\text{des}} = 0.6$).

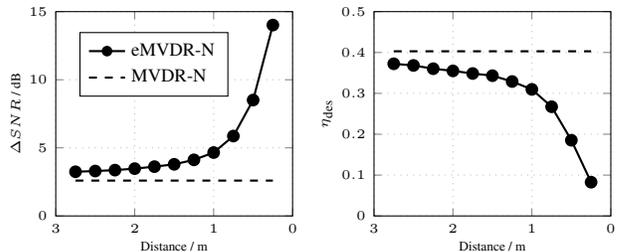


Fig. 3: SNR improvement (left) and optimal scaling parameter (right) for the MVDR-N and eMVDR-N beamformers for different distances between the external microphone and the speech source.

acoustically motivated [23] and don't alter the listener's impression of a spherically isotropic noise field.

For the MVDR-N and the eMVDR-N beamformer Fig. 3 depicts the SNR improvement and the optimal scaling parameter (both averaged over frequency) as a function of the distance between the external microphone and the speech source. As can be observed, the optimal scaling parameter resulting in the same psycho-acoustically motivated interaural noise coherence IC_n^{des} is smaller for the eMVDR-N beamformer than for the MVDR-N beamformer, hence leading to a larger SNR improvement. Therefore, for a homogeneous noise field the eMVDR-N beamformer provides a better trade-off between binaural cue preservation of the noise component and noise reduction performance than the MVDR-N beamformer, even for a realistic simulation. For the considered scenario the improvements are quite substantial when the distance between the external microphone and the speech source is smaller than 1 m.

6. CONCLUSION

In this paper we have shown both analytically and using simulations that for a homogeneous noise field the performance of the MVDR-N beamformer in terms of output SNR and preservation of the noise IC can be improved by incorporating an external microphone. The performance improvement is highly related to the position of the external microphone. These effects were also clearly observable in a simulation taking reverberation and estimation errors into account.

7. REFERENCES

- [1] V. Hamacher, U. Kornagel, T. Lotter, and H. Puder, "Binaural signal processing in hearing aids: Technologies and algorithms," in *Advances in Digital Speech Transmission*, pp. 401–429. Wiley, New York, NY, USA, 2008.
- [2] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, pp. 269–302. Wiley, 2010.
- [3] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel Signal Enhancement Algorithms for Assisted Listening Devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, Mar. 2015.
- [4] B. Cornelis, S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural multi-microphone noise reduction techniques," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 2, pp. 342–355, Feb. 2010.
- [5] T. Klasen, T. van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.
- [6] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural Coherence Preservation in MWF-based Binaural Noise Reduction Algorithms using Partial Noise Estimation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 654–658.
- [7] S. Markovich-Golan, A. Bertrand, M. Moonen, and S. Gannot, "Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks," *Signal Processing*, vol. 107, pp. 4–20, Feb. 2015.
- [8] T. C. Lawin-Ore, S. Stenzel, J. Freudenberger, and S. Doclo, "Alternative formulation and robustness analysis of the multichannel Wiener filter for spatially distributed microphones," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Juan les Pins, France, Sep. 2014.
- [9] A. Bertrand and M. Moonen, "Robust Distributed Noise Reduction in Hearing Aids with External Acoustic Sensor Nodes," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 14 pages, Jan. 2009.
- [10] N. Cvijanovic, O. Sadiq, and S. Srinivasan, "Speech enhancement using a remote wireless microphone," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 1, pp. 167–174, Feb. 2013.
- [11] D. Yee, H. Kamkar-Parsi, H. Puder, and R. Martin, "A speech enhancement system using binaural hearing aids and an external microphone," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 246–250.
- [12] J. Szurley, A. Bertrand, B. Van Dijk, and M. Moonen, "Binaural noise cue preservation in a binaural noise reduction system with a remote microphone signal," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 5, pp. 952–966, May 2016.
- [13] D. Marquardt, *Development and Evaluation of Psychoacoustically Motivated Binaural Noise Reduction and Cue Preservation Techniques*, Ph.D. thesis, Carl von Ossietzky Universität Oldenburg, 2015.
- [14] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical Analysis of Binaural Transfer Function MVDR Beamformers with Interference Cue Preservation Constraints," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.
- [15] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
- [16] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural Coherence Preservation in Multi-channel Wiener Filtering Based Noise Reduction for Binaural Hearing Aids," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2162–2176, Dec. 2015.
- [17] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [18] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [19] D. P. Jarret, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: Algorithm and application," *Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1462–1472, Sep. 2012.
- [20] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating non-stationary multisensor signals under a spatial coherence constraint," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.
- [21] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 22, no. 4, pp. 785–799, Apr. 2014.
- [22] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.
- [23] A. Walther and C. Faller, "Interaural correlation discrimination from diffuse field reference correlations," *Journal of the Acoustical Society of America*, vol. 133, no. 3, pp. 1496–1502, Mar. 2013.