Signal-Dependent Mixing for Direction-Preserving Multichannel Noise Reduction

Adrian Herzog
*International Audio Laboratories Erlangen**
91058 Erlangen, Germany
adrian.herzog@audiolabs-erlangen.de

Emanuël A.P. Habets
*International Audio Laboratories Erlangen**
91058 Erlangen, Germany
emanuel.habets@audiolabs-erlangen.de

***Abstract*—Reducing undesired sounds in a multichannel recording while preserving the spatial characteristics of the acoustic scene is a challenging task. In recent works by the present authors, different Ambisonics-to-Ambisonics noise reduction methods were discussed. To mitigate spatial distortions of the noise, a direction-preserving approach and a partial-noise-reduction approach were investigated. In this work, a signal-dependent partial-noise-reduction approach is proposed, which is able to mitigate spatial distortions of the noise further without significantly deteriorating other performance measures and can be applied to any multichannel signal format. The proposed method is evaluated using spherical-microphone-array recordings from the ACE corpus.**

***Index Terms*—Noise reduction, spatial distortions, Ambisonics**

I. Introduction

Multichannel acoustic signal enhancement is of paramount importance for, e.g., audio communication, hearing-aids or smart devices. Most existing techniques such as beamforming [1], [2] or signal separation [3], [4] aim at estimating the spectro-temporal characteristics of the desired sound component but not its spatial characteristics. For spatial sound acquisition, however, the spatial characteristics of the sound need to be preserved after signal enhancement. It has been shown, in the context of hearing aids, that it is crucial to not only preserve the spatial characteristics of the desired sound component but also of the undesired component, as the loss of spatial separation between these two sound components can deteriorate the intelligibility [5].

For general multichannel-to-multichannel noise reduction, different methods can be used which preserve the spatial characteristics of the desired component. The multidimensional Wiener filter (MWF) can be used for speech enhancement [6] or signal separation [7]. In the context of binaural beamforming, several methods have been proposed to preserve the spatial cues of the desired and undesired sound components [8], [9]. In [8], a partial noise reduction method was proposed which mixes the noise-reduced signal with the unprocessed signal. In previous works of the present authors [10], [11], Ambisonics-to-Ambisonics noise reduction was investigated. A direction-preserving approach was proposed and compared to the MWF with and without partial noise reduction. It was found, that the direction-preserving approach is better at preserving the spatial characteristics of the noise and computationally more efficient but can yield higher distortions of the desired sound compared to the MWF-based methods. Moreover, the direction-preserving approach was specifically formulated for Ambisonic signals. Therefore, it remains to develop an effective direction-preserving noise reduction method which does not increase the desired-signal distortion and works for arbitrary multichannel recordings.

In this work, we propose a new multichannel-to-multichannel noise reduction method which is derived by adding a measure for the spatial distortion of the noise to the MWF cost function. The result can be expressed as a MWF with signal-dependent mixing between the noise-reduced and unprocessed signals. We evaluate the proposed method using measured spherical-microphone-array impulse responses and noise signals from the ACE corpus [12] and show that the proposed method can be better at preserving the spatial distribution of the noise compared to the other methods without significantly deteriorating other performance measures.

II. Signal Model

A. General Signal Model

Let $\mathbf{x}$ denote a multichannel audio signal vector consisting of C channels. We assume that the signals are given in the short-time Fourier transform (STFT) domain, where the time-frame and frequency-bin indices are denoted by n and k, respectively. Moreover, we assume that $\mathbf{x}$ contains a desired component $\mathbf{s}$ and an undesired noise component $\mathbf{v}$, i.e.,

$$\mathbf{x}(k,n) = \mathbf{s}(k,n) + \mathbf{v}(k,n) \ . \tag{1}$$

Assuming that the desired component and the noise are mutually uncorrelated, the power spectral density (PSD) matrix

$$\mathbf{\Phi}_{\mathbf{x}}(k,n) = \mathcal{E}\left\{\mathbf{x}(k,n)\,\mathbf{x}^H(k,n)\right\} \ , \tag{2}$$

where $\mathcal{E}\{\cdot\}$ denotes the statistical expectation and $(\cdot)^H$ the Hermitian transpose, can be modeled as follows:

$$\mathbf{\Phi}_{\mathbf{x}}(k,n) = \mathbf{\Phi}_{\mathbf{s}}(k,n) + \mathbf{\Phi}_{\mathbf{v}}(k,n) \ , \tag{3}$$

where $\mathbf{\Phi}_{\mathbf{s}}$ and $\mathbf{\Phi}_{\mathbf{v}}$ denote the desired-signal and noise PSD matrices, respectively.

*A joint institution of the Friedrich Alexander University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

In this work, we consider a point source within a reverberant room for the desired component. In this case, $\mathbf{\Phi_s}$ can be decomposed into a direct-plus-early-reverberant part and a late-reverberant part as follows [13]:

$$\mathbf{\Phi_s}(k,n) = \phi_S(k,n)\,\mathbf{e}(k)\,\mathbf{e}^H(k) + \phi_R(k,n)\,\mathbf{\Gamma}_{\mathrm{dif}}(k)\ , \quad (4)$$

where ϕ_S, ϕ_R, $\mathbf{e}$ and $\mathbf{\Gamma}_{\mathrm{dif}}$ denote the direct-plus-early-reverberant PSD, the late-reverberant PSD, the relative-early transfer function (RETF) vector [13] and the diffuse spatial coherence matrix, respectively. For the noise, we assume a time-invariant spatial coherence matrix $\mathbf{\Gamma_v}$. Hence, $\mathbf{\Phi_v}$ is modeled as:

$$\mathbf{\Phi_v}(k,n) = \phi_V(k,n)\,\mathbf{\Gamma_v}(k) \quad (5)$$

with ϕ_V denoting the PSD of the noise.

B. Ambisonic Signal Model

Ambisonics [14], [15] is a full-sphere surround-sound format which represents the sound field via its spherical harmonic coefficients. Therefore, the C signals do not represent loudspeaker or microphone signals but different directional components of the sound field. Ambisonic signals can be obtained using spherical microphone arrays such as, e.g., the Eigenmike [16]. In this work, we consider Ambisonics as an exemplary mutichannel audio format to evaluate the proposed noise reduction method.

In Ambisonics, the signal channels are denoted by two indices $l = 0, ..., L$ and $m = -l, ...l$, denoted as order and degree, respectively. The maximum order L constitutes the Ambisonics order and the number of channels is $C = (L+1)^2$. The general signal model introduced in Sec. II-A can be used, where the diffuse spatial coherence matrix $\mathbf{\Gamma}_{\mathrm{dif}}$ becomes an identity matrix in the Ambisonics domain [17].

III. Noise Reduction using a Spatial Filter Matrix

We aim to reduce the undesired noise component of the noisy signal vector $\mathbf{x}$, while preserving the desired component $\mathbf{s}$. This can be achieved by applying a filter matrix $\mathbf{W}$ to $\mathbf{x}$ [10]:

$$\mathbf{z}(k,n) = \mathbf{W}(k,n)\,\mathbf{x}(k,n)\ , \quad (6)$$

where $\mathbf{z}$ is the C-dimensional multichannel signal vector with reduced noise. The STFT indices n and k are omitted for brevity in the remainder of this section.

A. Parametric Multidimensional Wiener Filter

The parametric multidimensional Wiener filter (PMWF) is the extension of the parametric multichannel Wiener filter [18] with multichannel output and can be derived, analogously to the MWF [6], by minimizing the cost function [11]

$$\mathcal{J}_\mu(\mathbf{W}) = \mathcal{E}\left\{\|\mathbf{W}\mathbf{s} - \mathbf{s}\|_2^2\right\} + \mu\,\mathcal{E}\left\{\|\mathbf{W}\mathbf{v}\|_2^2\right\} \quad (7)$$

w.r.t. $\mathbf{W}$, where $\|\cdot\|_2$ denotes the ℓ_2-norm and μ the trade-off parameter between speech distortion and noise reduction. The optimal solution $\mathbf{W}_\mu$ for is given by [6]

$$\mathbf{W}_\mu = \mathbf{\Phi_s}\left(\mathbf{\Phi_s} + \mu\,\mathbf{\Phi_v}\right)^{-1} \quad (8)$$

The ordinary MWF is obtained for $\mu = 1$. The PMWF preserves the spatial characteristics of the desired component well but distorts the spatial characteristics of the noise [11].

B. Partial Noise Reduction

To mitigate the spatial distortion of the noise, it has been proposed in [8] to only partially reduce the noise in the context of binaural beamforming. This idea can be formulated for general multichannel-to-multichannel noise reduction via the cost function [19]

$$\mathcal{J}_{a,\mu}(\mathbf{W}) = \mathcal{E}\left\{\|\mathbf{W}\mathbf{s} - \mathbf{s}\|_2^2\right\} + \mu\,\mathcal{E}\left\{\|\mathbf{W}\mathbf{v} - a\,\mathbf{v}\|_2^2\right\}\ , \quad (9)$$

where $a \in [0,1]$ denotes the mixing factor in this work. The optimal solution $\mathbf{W}_{a,\mu}$ can be expressed as follows:

$$\mathbf{W}_{a,\mu} = (1-a)\mathbf{W}_\mu + a\,\mathbf{I}\ , \quad (10)$$

where $\mathbf{I}$ denotes the $C \times C$-dimensional identity matrix. The partial noise reduction reduces the amount of spatial distortions of the noise at the cost of less noise reduction [19]. For $a = 0$, the PMWF $\mathbf{W}_\mu$ is obtained and, for $a = 1$, the filter matrix becomes an identity matrix, which yields no spatial distortions and no noise reduction.

C. Direction-Preserving Noise Reduction

The direction-preserving PMWF [10] is an Ambisonics-to-Ambisonics noise reduction method which utilizes a special form of the filter matrix $\mathbf{W}$ that preserves the directional information of any sound field in an optimal way.

The direction-preserving filter matrix can be expressed as follows [10]:

$$\mathbf{W}_{\mathrm{DP}} = \sum\nolimits_{q=1}^{Q} \alpha_q\,\mathfrak{q}_q \mathbf{y}^*(\Omega_q)\mathbf{y}^T(\Omega_q)\ , \quad (11)$$

where $\Omega_1, ..., \Omega_Q$ denote Q virtual spatial-sampling directions, $\mathfrak{q}_1, ..., \mathfrak{q}_Q$ the corresponding sampling weights, $\mathbf{y}(\Omega)$ the vector of spherical harmonic functions up to order L and $\alpha_1, ..., \alpha_Q$ are Q directional gain parameters. The symbols $(\cdot)^*$ and $(\cdot)^T$ denote the complex conjugate and transpose, respectively. The choice of the virtual sampling directions is discussed in more detail in [11].

Inserting $\mathbf{W}_{\mathrm{DP}}$ into the PMWF cost function (7), one can derive an expression for the directional gains α_q. The exact solution, however, requires a $Q \times Q$ matrix inversion as discussed in [11]. Therefore, an approximation was proposed in [10], [11], viz. $\mathbf{y}^T(\Omega_q)\mathbf{y}^*(\Omega_{q'}) \approx 0$ for $q \neq q'$, resulting in the following simplified expression for the directional gains [10]:

$$\alpha_q = \max\left\{\frac{\mathbf{y}^T(\Omega_q)\,\mathbf{\Phi_s}\,\mathbf{y}^*(\Omega_q)}{\mathbf{y}^T(\Omega_q)\,\left(\mathbf{\Phi_s} + \mu\,\mathbf{\Phi_v}\right)\,\mathbf{y}^*(\Omega_q)}, \alpha_{\min}\right\}\ , \quad (12)$$

where a lower bound $\alpha_{\min} \in [0,1)$ for the gains α_q has been introduced to reduce audible distortions. The direction-preserving PMWF yields less spatial distortion of the noise compared to the PMWF with partial noise reduction (10), at the cost of higher desired-signal distortion [11]. Moreover, it can be computationally more efficient compared to the PMWF (8) which requires a $C \times C$ matrix inversion.

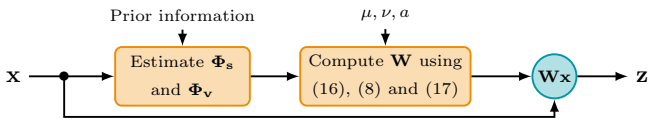

Fig. 1. Block diagram of proposed noise reduction method.

IV. Proposed Noise Reduction Method

In our previous works, we used the direction-preserving form of the filter matrix (11) to mitigate spatial distortions of the noise. In this work, we do not constrain the filter matrix, but add a measure for the spatial distortion of the noise to the cost function (9). For this purpose, we compute how much the residual noise $\mathbf{Wv}$ differs from a scaled version of the noise given by $b\,\mathbf{v}$. In contrast to the mixing parameter a, the scalar b is not a free parameter but chosen such that the expected distance between $\mathbf{Wv}$ and $b\,\mathbf{v}$ is minimal. This results in the following measure for the spatial distortion of the noise:

$$\begin{aligned}\mathcal{J}_{\mathrm{ND}}(\mathbf{W}) &= \min_b \mathcal{E}\left\{\|\mathbf{Wv} - b\,\mathbf{v}\|_2^2\right\} \\ &= \mathcal{E}\left\{\left\|\mathbf{Wv} - \frac{\mathrm{tr}\{\mathbf{W\Phi_v}\}}{\mathrm{tr}\{\mathbf{\Phi_v}\}}\,\mathbf{v}\right\|_2^2\right\} , \end{aligned} \tag{13}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace operator. The compound cost function becomes:

$$\mathcal{J}_{a,\mu\,\nu}(\mathbf{W}) = \mathcal{J}_{a,\mu}(\mathbf{W}) + \nu\,\mathcal{J}_{\mathrm{ND}}(\mathbf{W}) , \tag{14}$$

where, in addition to a, ν is a trade-off parameter between noise reduction and spatial distortion of the noise.

Computing the gradient of (14) w.r.t. $\mathbf{W}^H$, yields the following set of equations:

$$\mathbf{W}\left[\mathbf{\Phi_s} + (\mu+\nu)\mathbf{\Phi_v}\right] - \nu\frac{\mathrm{tr}\{\mathbf{W\Phi_v}\}}{\mathrm{tr}\{\mathbf{\Phi_v}\}}\mathbf{\Phi_v} = \mathbf{\Phi_s} + a\mu\,\mathbf{\Phi_v} . \tag{15}$$

This linear system of equations can be solved using the ansatz:

$$\mathbf{W} = (1-a')\,\mathbf{W}_{\mu+\nu} + a'\mathbf{I} , \tag{16}$$

where $\mathbf{W}_{\mu+\nu}$ is defined analogously to (8) and a' is a scalar. Inserting (16) into (15) and solving for a' yields:

$$a' = a + (1-a)\frac{\nu\,\mathrm{tr}\{\mathbf{W}_{\mu+\nu}\mathbf{\Phi_v}\}}{\mu\,\mathrm{tr}\{\mathbf{\Phi_v}\} + \nu\,\mathrm{tr}\{\mathbf{W}_{\mu+\nu}\mathbf{\Phi_v}\}} . \tag{17}$$

Hence, the proposed noise reduction method can be interpreted as a PMWF with signal-dependent mixing factor a', where a is the lower bound of a'. For $\nu = 0$, the PMWF with signal-dependent mixing reduces to the PMWF $\mathbf{W}_{a,\mu}$ with fixed mixing factor a. For $\mu = 0$, the mixing factor becomes $a' = 1$ for all ν and, therefore, the filter matrix becomes an identity matrix. The processing scheme of the proposed method is shown in Fig. 1.

V. Parameter Estimation

The discussed noise reduction methods require estimates of the desired and noise PSD matrices $\mathbf{\Phi_s}$ and $\mathbf{\Phi_v}$. Using the signal model for the desired component (4) and the noise (5), we have to estimate the early- and late-reverberant PSDs ϕ_S and ϕ_R, the noise PSD ϕ_V, the RETF vector $\mathbf{e}$ and the spatial coherence matrix of the noise $\mathbf{\Gamma_v}$.

We estimate these parameters using the method in [11], which can be summarized as follows:

- *Prior information:* Assume knowledge of $\mathbf{\Gamma_v}$ and the reverbereration time T_{60}. In this work, we computed $\mathbf{\Gamma_v}$ by averaging $\mathbf{vv}^H$ over all time frames and normalizing w.r.t. the first element. In practice, $\mathbf{\Gamma_v}$ can be estimated when the desired sound is inactive. The reverberation time is needed for the PSD estimation.
- Offline estimation of $\mathbf{e}$ using the covariance-whitening method [20] and the direct-to-reverberant ratio κ as described in [11], where κ is required for the subsequent PSD estimation.
- Recursive estimation of the noisy PSD matrix via:
$$\hat{\mathbf{\Phi}}_{\mathbf{x}}(k,n) = \beta\,\hat{\mathbf{\Phi}}_{\mathbf{x}}(k,n-1) + (1-\beta)\,\mathbf{x}(k,n)\,\mathbf{x}^H(k,n) ,$$
where $\beta \in [0,1)$ is a recursive smoothing parameter.
- Recursive estimation of ϕ_R using T_{60}, $\hat{\kappa}$ and the late-reverberant PSD model from [21].
- Estimation of ϕ_S and ϕ_V using $\hat{\mathbf{\Phi}}_{\mathbf{x}}$, $\hat{\phi}_R$, $\hat{\mathbf{e}}$ and the Frobenius-norm PSD estimator [22].

Estimated quantities have been denoted with the hat symbol $\hat{\cdot}$. See [11], for more details on the parameter estimation.

VI. Evaluation[1]

A. Performance Measures

To evaluate the performance of the discussed noise reduction methods, we computed the following noise reduction (NR) and signal-to-distortion ratio (SDR) measures per time-segment t and Mel-scale [23] frequency band b:

$$\begin{aligned}\mathrm{SDR}(t,b) &= 10\log_{10}\left(\frac{\sum_{n\in\mathcal{N}_t,k\in\mathcal{K}_b}\|\mathbf{s}(k,n)\|_2^2}{\sum_{n\in\mathcal{N}_t,k\in\mathcal{K}_b}\|[\mathbf{Ws}-\mathbf{s}](k,n)\|_2^2}\right) \\ \mathrm{NR}(t,b) &= 10\log_{10}\left(\frac{\sum_{n\in\mathcal{N}_t,k\in\mathcal{K}_b}\|\mathbf{v}(k,n)\|_2^2}{\sum_{n\in\mathcal{N}_t,k\in\mathcal{K}_b}\|[\mathbf{Wv}](k,n)\|_2^2}\right) ,\end{aligned} \tag{18}$$

where the time-segments and frequency bands are defined via

$$\begin{aligned}\mathcal{N}_t &= \{(t-1)T+1,(t-1)T+2,...,tT\} \quad \text{with } T=6 \\ \mathcal{K}_b &= \{k_{b-1}+1,k_{b-1}+1,...,k_b\} \quad \text{with} \\ k_b &= \left\lfloor 700\left(\left(1+\frac{f_s/2}{700}\right)^{b/B}-1\right)\frac{N_{\mathrm{DFT}}}{f_s}\right\rfloor\end{aligned} \tag{19}$$

with $B = 13$ denoting the number of Mel-scale bands, f_s the sampling frequency and N_{DFT} the discrete Fourier transform (DFT) size of the STFT. The choice of the time-segment length and frequency bands was motivated by commonly used objective quality measures such as the frequency-weighted segmental signal-to-noise ratio [24]. The Mel-scale is a perceptually-motivated frequency scale based on equally-perceived pitch differences [23].

[1]Audio examples are provided at https://www.audiolabs-erlangen.de/resources/2021-EUSIPCO-Partial-Noise-Reduction

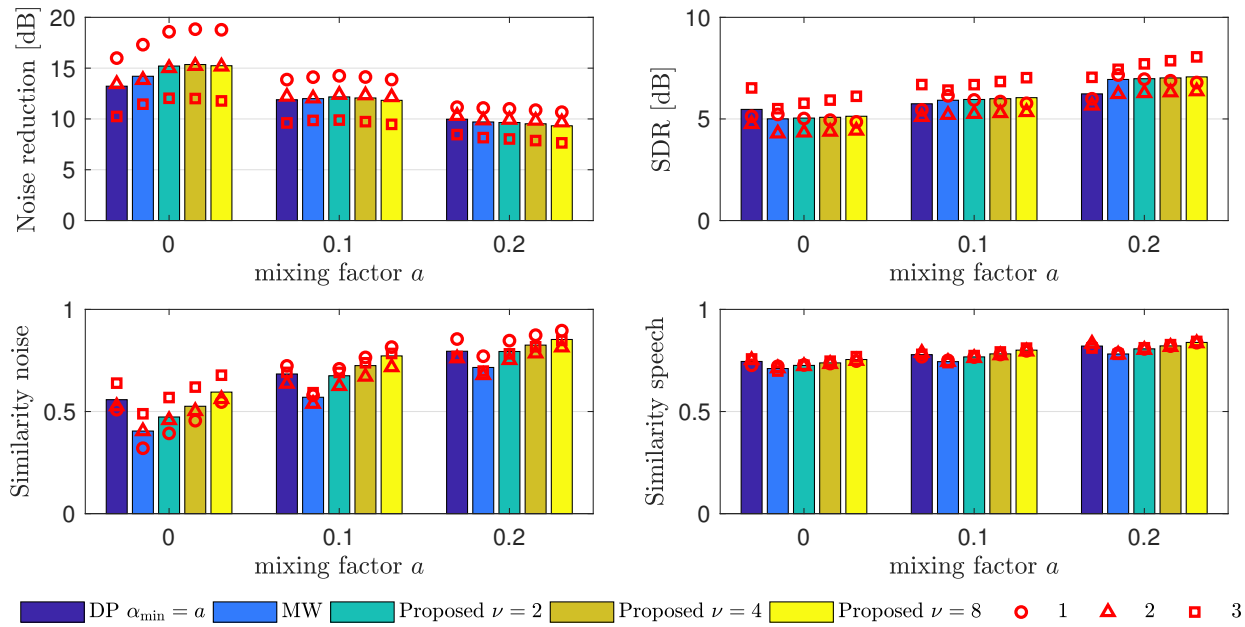

Fig. 2. Performance measures for different mixing factors a and noise reduction methods averaged over all configurations.

To evaluate spatial distortions, we used the following similarity measure [11]:

$$\sigma_{\mathbf{p},\mathbf{q}}(t,b) = \frac{2}{\pi}\sin^{-1}\left(\frac{\operatorname{tr}\left\{\bar{\mathbf{\Phi}}_{\mathbf{p}}(t,b)\bar{\mathbf{\Phi}}_{\mathbf{q}}(t,b)\right\}}{\sqrt{\operatorname{tr}\left\{\bar{\mathbf{\Phi}}^2_{\mathbf{p}}(t,b)\right\}\operatorname{tr}\left\{\bar{\mathbf{\Phi}}^2_{\mathbf{q}}(t,b)\right\}}}\right) \tag{20}$$

with $(\mathbf{p},\mathbf{q}) = (\mathbf{v},\mathbf{W}\mathbf{v})$ (similarity noise) or $(\mathbf{p},\mathbf{q}) = (\mathbf{s},\mathbf{W}\mathbf{s})$ (similarity speech) and

$$\bar{\mathbf{\Phi}}_{\mathbf{p}}(t,b) = \sum\nolimits_{n\in\mathcal{N}_t, k\in\mathcal{K}_b} \mathbf{p}(k,n)\mathbf{p}^H(k,n)\ . \tag{21}$$

Note, that $\sigma_{\mathbf{p},\mathbf{q}}$ measures the angular similarity between the two segmental PSD matrices $\bar{\mathbf{\Phi}}_{\mathbf{p}}$ and $\bar{\mathbf{\Phi}}_{\mathbf{q}}$. A similarity of $\sigma_{\mathbf{p},\mathbf{q}} = 1$ indicates that the spatial distributions of the signal vectors $\mathbf{p}$ and $\mathbf{q}$ are the same, whereas $\sigma_{\mathbf{p},\mathbf{q}} = 0$ indicates very different spatial distributions.

For the evaluation, the NR and SDR measures were limited to the range $[-10, 35]$ dB as suggested in [24] and, subsequently, all measures were averaged over the segments and frequency bands.

B. Evaluation Setup

Three different noisy and reverberant Ambisonic signals with $L = 3$ were generated for the evaluation. Different English speech files with $f_s = 16$ kHz and a length of 5 seconds were used for the source signals. For the reverberation and noise, Eigenmike [16] recordings from the ACE corpus [12] were used. For each of the three Ambisonic signals, a speech file, a room and a noise type were selected. The speech file was upsampled to 48 kHz (the sampling rate of the Eigenmike recordings) and convolved with the Eigenmike room-impulse responses of the selected room. Next, the Eigenmike noise signal corresponding to the selected room and noise type was scaled according to the desired signal to noise ratio (SNR). The Ambisonic signals for the desired speech component and the noise were obtained using the EigenUnits software [25].

For the processing, the signal components were downsampled to 16 kHz and transformed to the STFT domain using a square-root Hann window of 512 samples (32 ms) length, 50% overlap and a DFT size of 1024. The mixture signals were obtained by adding the speech and noise signals with an SNR of 0 dB.

The following configurations were used:

1) Female speech in "Office 2" with $T_{60} = 0.39$ s and babble noise (○).
2) Female speech in "Lecture Room 1" with $T_{60} = 0.64$ s and fan noise (△).
3) Male speech in "Lecture Room 2" with $T_{60} = 1.25$ s and ambient noise (□).

For the noise reduction, we chose $\beta = 0.5$ and $\mu = 1$. For the direction-preserving method, a lower-bound $\alpha_{\min}$ equal to the mixing factor a and $Q = 49$ almost uniformly distributed virtual sampling directions were chosen as in [11]. The matrix inversion of the matrix PMWF in (8) was regularized by adding $10^{-9}\,\mathbf{I}$ to the matrix before inversion.

C. Results

In Fig. 2, the performance measures of the direction-preserving PMWF (DP), the PMWF with partial noise reduction (MW) and the proposed noise reduction method (Proposed) are shown for different mixing factors a and noise-distortion trade-off parameters ν. Mean results are represented by bars and individual results for the configurations 1 – 3 with the symbols ○, △ and □, respectively.

For all methods, the NR decreased with increasing a while the other performance measures increased with a. In contrast, increasing the trade-off parameter ν for the proposed method, did not significantly influence the NR and SDR results while increasing the similarity measures. For $a = 0$, the DP method yielded the lowest NR but the best SDR, which reflects the noise-reduction vs. speech distortion trade-off of the PMWF. For $a = 0.1$, all methods resulted in very similar NR and SDR measures while, for $a = 0.2$ the DP method yielded slightly lower SDRs. The best NR results were obtained for configuration 1 (○) which has the lowest T_{60} making the desired reverberant sound less diffuse and thus easier to separate from the noise. On the contrary, configuration 3 (□) with the largest T_{60} yielded the lowest NR results. Note, that the NR and SDR performances are limited by the assumed reverberant signal model which is not exactly fulfilled for the investigated Ambisonic signals.

The differences in noise similarity for the different methods are most prominent. As discussed in [11], the DP method is better at preserving the spatial distribution of the noise compared to the MW method. The noise similarity of the proposed method increased with increasing ν while the NR performance decreased only slightly and the SDR increased slightly. The proposed method can achieve higher noise similarities than the DP method when ν is large enough without significantly deteriorating the other performance measures. The speech similarity measures were high for all methods, slightly increased for the proposed method with increasing ν and became comparable to the speech similarity measures of the DP method for $\nu \in \{4, 8\}$.

VII. Conclusion

We proposed a new method to mitigate spatial distortions of the noise for the PMWF. The method can be expressed as a PMWF with partial noise reduction and signal-dependent mixing factor. In the evaluation, we found that the proposed noise reduction method can achieve higher noise similarities than all other methods without significantly deteriorating the other performance measures. It should be noted that the DP method does not require a matrix inversion, while the proposed method does due to the PMWF matrix. Therefore, the authors recommend to use the proposed method whenever the computational complexity is feasible and the DP method when the computational complexity needs to be reduced.

References

[1] B. D. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoustics, Speech and Signal Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer-Verlag, Berlin, Germany, 2008.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley-Interscience, 2001.

[4] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures: A unified treatment," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, J. Huang and J. Benesty, Eds., pp. 255–293. Kluwer Academic Publ., Boston, Feb. 2004.

[5] J. Peissig and B. Kollmeier, "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.*, vol. 101, no. 3, pp. 1660–1670, 1997.

[6] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sept. 2002.

[7] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1830–1840, May 2010.

[8] T. J. Klasen, T. V. den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, 2007.

[9] T.V. den Bogaert, J. Wouters, S. Doclo, and M. Moonen, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel wiener filter," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, April 2007, vol. 4, pp. IV–565–IV–568.

[10] A. Herzog and E. A. P. Habets, "Direction-preserving Wiener matrix filtering for Ambisonic input-output systems," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

[11] A. Herzog and E. A. P. Habets, "Direction and reverberation preserving noise reduction of ambisonics signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 2461–2475, Aug. 2020.

[12] J. Eaton, N. D. Gaubich, A. H. Moore, and P.A. Naylor, "The ACE challenge - corpus description and performance evaluation," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2015.

[13] O. Schwartz, S. Gannot, and E.A.P. Habets, "Multi-microphone speech dereverberation and noise reduction using relative early transfer functions," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 2, pp. 240–251, Jan. 2015.

[14] M. A. Gerzon, "Periphony: With-height sound reproduction," *Journal Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, Feb. 1973.

[15] F. Zotter and M. Frank, *Ambisonics - A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer, 2019.

[16] mh acoustics, LLC, "em32 Eigenmike," .

[17] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*, Springer, 2017.

[18] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.

[19] B. Cornelis, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[20] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[21] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 770–774, Sept. 2009.

[22] O. Schwartz, S. Gannot, and E. A. P. Habets, "Joint estimation of late reverberant and speech power spectral densities in noisy environments using Frobenius norm," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aug. 2016, pp. 1123–1127.

[23] Stanley Smith Stevens, John Volkman, and Edwin B. Newman, "A scale for the measurement of the psychological magnitude of pitch," *J. Acoust. Soc. Am.*, vol. 8, pp. 185–19, 1937.

[24] Y. Hu and P.C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[25] mh acoustics, LLC, *EigenUnits user manual version 2 rev. B*, Oct. 2019.