KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Arnout Roebben, Toon van Waterschoot, Jan Wouters, and Marc Moonen Title: Integrated Minimum Mean Squared Error Algorithms for Combined Acoustic Echo Cancellation, Dereverberation, and Noise Reduction
Archived version	■ Submitted version ☐ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ☐ EU H22020 ☐ EU Horizon Europe ■ FWO ■ KULeuven Internal Funds ☐ VLAIO ☐ AI ☐ Other



Integrated Minimum Mean Squared Error Algorithms for Combined Acoustic Echo Cancellation, Dereverberation, and Noise Reduction

Arnout Roebben, *Graduate Student Member, IEEE*, Toon van Waterschoot, *Member, IEEE*, Jan Wouters, and Marc Moonen, *Fellow, IEEE*

***Abstract*—In various speech recording applications, the direct and early reverberation component of the desired speech signal is corrupted by a late reverberation component, as well as by acoustic echo and noise. Hence, algorithms for combined acoustic echo cancellation (AEC), dereverberation (DR), and noise reduction (NR) are required. In principle, plainly cascading an AEC, DR, and NR algorithm in any particular order ignores their interaction, such that in this paper, an integrated algorithm design approach is proposed. A general multi-microphone/multi-loudspeaker scenario with possible linear dependencies between (delayed) microphone and loudspeaker signals is considered. First, an extended short-time Fourier transform (STFT) domain signal model is formulated, which comprises the microphone and loudspeaker signals, and their delayed versions. Second, a common minimum mean squared error (MSE) cost function is formulated, which encompasses a weighted combination of residual echo, late reverberation and noise, and the distortion of the direct and early reverberation component of the desired speech signal. Third, using a common solution strategy, it is shown that the resulting speech distortion weighted multi-channel Wiener filter (SD-MWF) on the extended signal model, referred to as SD-MWF$_{\text{ext}}$, is not implementable. Nevertheless, this SD-MWF$_{\text{ext}}$ is shown to be equivalent to a specific cascade of AEC, DR, and NR filters. The weighting, introduced in the cost function, reappears in the NR filter. Under rank-deficiency conditions, the SD-MWF$_{\text{ext}}$ is non-unique. This theoretical equivalence then amounts to the cascade being a specific, not necessarily minimum-norm equivalent solution for this SD-MWF$_{\text{ext}}$. The cascaded implementation is experimentally shown to be effective.**

***Index Terms*—Integrated algorithm design approach, Audio signal processing, Multi-channel, Acoustic echo cancellation (AEC), Dereverberation (DR), and Noise reduction (NR).**

I. Introduction

In several speech recording applications, e.g., hands-free telephony in cars and hearables, the recorded microphone signal comprises a desired speech signal together with undesired acoustic echo and noise. The desired speech and noise originate from sources within the room (the near-end room) [1]. The acoustic echo originates from loudspeakers located within the near-end room, playing back signals recorded in another room (the far-end room) [1]. Similarly to the noise, the recorded desired speech signal can contain the contribution of multiple desired speech sources. Additionally, this recorded desired speech signal can be decomposed into a direct and early reverberation component, both contributing to speech intelligibility, and a late reverberation component, degrading speech intelligibility [2]. To only retain this direct and early reverberation component, while suppressing the echo, late reverberation, and noise, an algorithm for combined acoustic echo cancellation (AEC), dereverberation (DR), and noise reduction (NR) is required.

AEC algorithms typically exploit the loudspeaker signals to model the echo path between the loudspeakers and the microphones, to estimate, and subsequently cancel the acoustic echo in the microphone signals [1], [3]. Residual echo is further suppressed by applying a filter on the microphone signals [1], [4]. DR algorithms can generally be divided into two classes, inverse-filtering-based [5]–[8] and beamforming-based algorithms [9], [10]. In inverse-filtering-based DR algorithms, delayed microphone signals are used to model the multi-microphone inverse of the acoustic transfer functions (ATFs) between the desired speech sources and the microphones. In beamforming-based DR algorithms, the late reverberation component is treated as originating from an isotropic, e.g., diffuse, sound field, such that NR-like algorithms can be applied. NR algorithms apply a filter to the microphone signals to suppress the noise [11]–[14]. These AEC, DR, and NR algorithms can be designed according to a model-based design approach, i.e., employing a signal model to derive the algorithm [1], [5], [6], [9]–[12]. A hybrid design approach can also be adopted. A neural network (NN) is then employed for parameter estimation [3], [7], [13], or is cascaded with a model-based algorithm [4], [8]. In a data-driven design approach, the NN directly estimates the direct and early reverberation component of the desired speech signal from the microphone signals [14], [15].

Algorithms that implement multiple functionalities, e.g., combinations of AEC, DR, and NR, can be designed according to a cascaded or an integrated approach [16]. In a cascaded approach, existing single-functionality algorithms are cascaded in any particular order, thereby ignoring the interaction between the functionalities. The parameters of these cascaded algorithms can nevertheless still be jointly optimised, e.g., using NNs or expectation maximisation (EM) [17]. In an integrated approach, on the other hand, this interaction is explicitly accounted for by designing a single algorithm that implements all functionalities and originates from a single cost function. For combined DR and NR, integrated filters

This research was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven project C14-21-0075 'A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices', and Aspirant Grant 11PDH24N (for A. Roebben) from the Research Foundation - Flanders (FWO). The scientific responsibility is assumed by its authors.

Arnout Roebben, Toon van Waterschoot and Marc Moonen are with the Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, B-3001 Leuven, Belgium (e-mail: arnout.roebben@esat.kuleuven.be; toon.vanwaterschoot@esat.kuleuven.be; marc.moonen@esat.kuleuven.be).

Jan Wouters is with the Department of Neurosciences, Research Group ExpORL, KU Leuven, B-3000 Leuven, Belgium (e-mail: jan.wouters@med.kuleuven.be).

have been proposed, e.g., multi-frame minimum power distortionless response (MPDR) beamformers [18], [19], single-frame multi-channel Wiener filters (MWFs) that are based on the assumption that the late reverberation originates from an isotropic sound field [9], or NNs that perform DR and NR simultaneously [15]. Yet, the multi-frame MPDR beamformer has been shown to be equivalent to a cascade of an inverse-filtering-based DR filter and a single-frame MPDR NR filter [19]. Alternative approaches, e.g., Kalman filters, also exist [20]. For combined AEC and NR, integrated filters have similarly been proposed in the form of extended linearly constrained minimum variance-based filters and extended MWFs that consider both microphone and loudspeaker signals [16], [21]–[23], as well as in the form of NNs that perform simultaneous AEC and NR [4]. Yet, the extended MWFs have been shown to be equivalent to a cascade, e.g., of AEC and NR filters [16], [22]. Again, alternative approaches, including Kalman filters, also exist [24]. Algorithm design for combined AEC, DR, and NR has only received limited attention, both using a cascaded approach [17], [25]–[28], and using an integrated approach. These integrated approaches focused mostly on state-space models or NNs, and target specific scenarios, e.g., with a single loudspeaker [29]–[33].

In this paper, an integrated algorithm design approach is proposed for the problem of combined AEC, DR, and NR. A general multi-microphone/multi-loudspeaker scenario with possible linear dependencies between (delayed) microphone and loudspeaker signals is considered. First, an extended short-time Fourier transform (STFT) domain signal model is formulated, which comprises the microphone and loudspeaker signals, and their delayed versions. Second, a common minimum mean squared error (MSE) cost function is formulated, which encompasses a weighted combination of residual echo, late reverberation and noise, and the distortion of the direct and early reverberation component of the desired speech signal. Third, using a common solution strategy, it is shown that the resulting speech distortion weighted multi-channel Wiener filter (SD-MWF) on the extended signal model, referred to as SD-MWF$_{\text{ext}}$, is not implementable. This SD-MWF$_{\text{ext}}$ is subsequently shown to be equivalent to a cascade of AEC, DR, and NR filters. Under rank-deficiency conditions, the SD-MWF$_{\text{ext}}$ is non-unique, such that this theoretical equivalence amounts to the cascade being a specific, not necessarily minimum-norm, equivalent solution for this SD-MWF$_{\text{ext}}$. This paper generalises [16], where a similar strategy has been shown effective for combined AEC and NR. The resulting cascade of AEC, DR, and NR filters is visualised in Fig. 1.

The signal model is described in Section II, and the cost function in Section III. The solution is derived and analysed theoretically in Section IV, and further analysed experimentally, according to the experimental design of Section V, in Section VI. Finally, Section VII draws the conclusions. MATLAB code is available in [35].

II. Signal Model

A. Microphone Signal Model

An M-microphone setup is considered. Microphone signal $m_i[f,k]$, $i \in \{1,\cdots,M\}$, specified in the STFT domain with

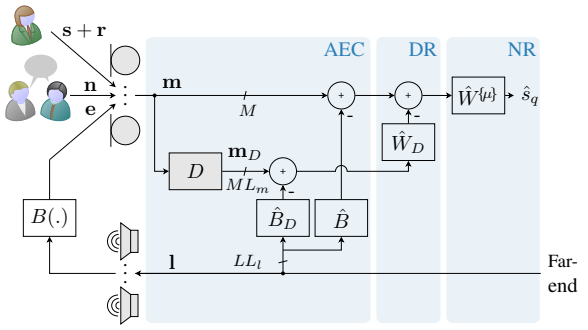

Fig. 1. Block diagram of the solution for combined acoustic echo cancellation (AEC), dereverberation (DR), and noise reduction (NR). The figurines (in purple) have been generated using [34].

frequency bin index $f \in \{0,\cdots,F-1\}$ and frame index $k \in \{0,\cdots,K-1\}$, is given as

$$m_i[f,k] = s_i[f,k] + r_i[f,k] + e_i[f,k] + n_i[f,k]. \quad (1)$$

In (1), $s_i[f,k]+r_i[f,k]$ represents the recorded reverberant desired speech signal in microphone i, with direct and early reverberation component $s_i[f,k]$ and late reverberation component $r_i[f,k]$. Similarly, $e_i[f,k]$ and $n_i[f,k]$ represent the echo and noise signals.

The recorded reverberant desired speech signal in microphone i, $s_i[f,k]+r_i[f,k]$, contains the contributions of N_s desired speech sources that each adhere to a convolutive transfer function (CTF) model [5], [36]. In this CTF model, the reverberant desired speech signal is assumed to be generated as the convolution of ATFs $h_{n,i}[f,k]$, $n = \{1,\cdots,N_s\}$, and source signals $d_n[f,k]$ across L_h frames as [5], [36]

$$\begin{aligned} s_i[f,k] + r_i[f,k] &= \sum_{n=1}^{N_s}\sum_{p=0}^{L_h-1} h_{n,i}[f,p]^* d_n[f,k-p] \quad (2)\\ &\triangleq \sum_{n=1}^{N_s}\sum_{p=0}^{D-1} h_{n,i}[f,p]^* d_n[f,k-p] + \sum_{n=1}^{N_s}\sum_{p=D}^{L_h-1} h_{n,i}[f,p]^* d_n[f,k-p]. \end{aligned}$$

In (2), the boundary between the early reverberation, and the late reverberation is set by frame index D.

The echo signal $e_i[f,k]$ originates from the loudspeaker signals $l_j[f,k]$, $j \in \{1,\cdots,L\}$. Let the map $B_i(.): \mathbb{C}^{LL_g} \to \mathbb{C}$ represent the echo path for microphone i, i.e., the propagation between the loudspeakers and microphone i [16]. The echo signal for microphone i is then generated by applying $B_i(.)$ to the lagged loudspeaker signals with L_g lags as

$$e_i[f,k] = B_i\Big(\big[l_1[f,k]\cdots l_1[f,k-L_g+1]\cdots l_L[f,k-L_g+1]\big]^{\mathsf{T}}\Big). \quad (3)$$

While this map $B_i(.)$ can be a general map, e.g., containing loudspeaker non-linearities, it can also be restricted to a linear map. For this linear map, (3) simplifies to [16]

$$e_i[f,k] = \mathbf{b}_i[f]^H\big[l_1[f,k]\cdots l_1[f,k-L_g+1]\cdots l_L[f,k-L_g+1]\big]^{\mathsf{T}}, \quad (4)$$

with $\mathbf{b}_i[f]$ containing the coefficients of the linear map as

$$\mathbf{b}_i[f] = \big[b_{i,1}[f,0]\cdots b_{i,1}[f,L_g-1]\cdots b_{i,L}[f,L_g-1]\big]^{\mathsf{T}}. \quad (5)$$

The microphone signals can be stacked into a microphone signal vector $\mathbf{m}[f,k] \in \mathbb{C}^M$ as

$$\mathbf{m}[f,k] = \begin{bmatrix} m_1[f,k] & \cdots & m_M[f,k] \end{bmatrix}^\mathsf{T} \tag{6a}$$
$$= \mathbf{s}[f,k] + \mathbf{r}[f,k] + \mathbf{e}[f,k] + \mathbf{n}[f,k], \tag{6b}$$

with similar definitions for $\mathbf{s}[f,k]$, $\mathbf{r}[f,k]$, $\mathbf{e}[f,k]$, and $\mathbf{n}[f,k]$. As $\mathbf{s}[f,k]+\mathbf{r}[f,k]$ represents the reverberant desired speech signal vector, it attains an on-off behaviour [37], [38]. A voice activity detector (VAD), $\text{VAD}[k]$, is used to detect these periods of reverberant desired speech activity [37], [38]. The echo and noise are assumed to be always active [1], [16], such that the following regimes can be defined

$$\begin{cases} \text{VAD}[k]=1 : \mathbf{m}[f,k]=\mathbf{s}[f,k]+\mathbf{r}[f,k]+\mathbf{e}[f,k]+\mathbf{n}[f,k] \text{ recorded} \\ \text{VAD}[k]=0 : \mathbf{e}[f,k]+\mathbf{n}[f,k] \text{ recorded.} \end{cases}$$

Thus, $\mathbf{s}[f,k]$ can never be observed separately from $\mathbf{r}[f,k]$.

B. Assumptions

The following assumptions are made:

- As common in DR literature, each source signal $d_n[f,k]$ is assumed to be temporally uncorrelated across frames [5], [20], [36], i.e., $\mathbb{E}\{d_n[f,k]d_n[f,k-p]\} = 0,\ \forall p \neq 0$, where $\mathbb{E}\{.\}$ is the expected value operator. For a sufficiently large frame and hop size, this assumption can be considered approximately justified. Thus, the direct and early reverberation component of the desired speech signal vector is uncorrelated with the late reverberant component, i.e, $\mathbb{E}\{\mathbf{s}[f,k]\mathbf{r}[f,k]^H\} = 0$, with 0 denoting the zero-matrix. [5], [9], [20].
- The reverberant desired speech $\mathbf{s}[f,k] + \mathbf{r}[f,k]$, echo $\mathbf{e}[f,k]$, and noise $\mathbf{n}[f,k]$ signal vectors are assumed to be uncorrelated as they originate from different sources.

Thus, all components in (6b) are assumed to be mutually uncorrelated. No rank-assumptions are imposed. Let $R_{yz}[f,k] = \mathbb{E}\{\mathbf{y}[f,k]\mathbf{z}[f,k]^H\}$ be the correlation matrix between general signal vectors $\mathbf{y}[f,k]$ and $\mathbf{z}[f,k]$. Then, $R_{mm}[f,k]$, $R_{ss}[f,k]$, $R_{ee}[f,k]$, $R_{nn}[f,k]$, and $R_{ll}[f,k]$ are defined accordingly. The correlation matrices can be rank-deficient, modelling localised desired speech, echo, and noise sources, as well as linearly related loudspeaker and microphone signals.

C. Extended Signal Model

An extended signal model can be constructed using $\mathbf{m}_D[f,k] \in \mathbb{C}^{ML_m}$, containing the lagged microphone signals from lag D (as defined in (2)) up to lag $D+L_m-1$, and using $\mathbf{l}[f,k] \in \mathbb{C}^{LL_l}$, containing the lagged loudspeaker signals up to lag L_l-1, i.e.,

$$\mathbf{m}_D[f,k] = \begin{bmatrix} m_1[f,k-D]\cdots m_1[f,k-D-L_m+1]\cdots m_M[f,k-D-L_m+1] \end{bmatrix}^\mathsf{T} \tag{7a}$$
$$\mathbf{l}[f,k] = \begin{bmatrix} l_1[f,k] & \cdots & l_1[f,k-L_l+1] & \cdots & l_L[f,k-L_l+1] \end{bmatrix}^\mathsf{T}. \tag{7b}$$

With this, $\tilde{\mathbf{m}}[f,k] \in \mathbb{C}^E$, with $E = M+ML_m+LL_l$, is defined as the extended signal vector, i.e.

$$\tilde{\mathbf{m}}[f,k] = \begin{bmatrix} \mathbf{m}[f,k] \\ \mathbf{m}_D[f,k] \\ \mathbf{l}[f,k] \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{m}[f,k] \\ \mathbf{t}[f,k] \end{bmatrix}. \tag{8}$$

Similar to (6b), $\tilde{\mathbf{m}}[f,k]$ can be decomposed into its individual components, i.e.,

$$\tilde{\mathbf{m}}[f,k] = \begin{bmatrix} \mathbf{s}[f,k] \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{r}[f,k] \\ \mathbf{s}_D[f,k] + \mathbf{r}_D[f,k] \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{e}[f,k] \\ \mathbf{e}_D[f,k] \\ \mathbf{l}[f,k] \end{bmatrix} + \begin{bmatrix} \mathbf{n}[f,k] \\ \mathbf{n}_D[f,k] \\ \mathbf{0} \end{bmatrix} \tag{9a}$$
$$= \tilde{\mathbf{s}}[f,k] + \tilde{\mathbf{r}}[f,k] + \tilde{\mathbf{e}}[f,k] + \tilde{\mathbf{n}}[f,k], \tag{9b}$$

with definitions for $\mathbf{s}_D[f,k]$, $\mathbf{r}_D[f,k]$, $\mathbf{e}_D[f,k]$ and $\mathbf{n}_D[f,k]$ similar to (7a), and $\mathbf{0}$ a zero-vector. Under the assumptions of Section II-B, the components in (9b) are mutually uncorrelated. The signal vector $\mathbf{s}_D[f,k]$ is stacked below $\mathbf{r}[f,k]$ as it is uncorrelated to $\mathbf{s}[f,k]$, yet, it is correlated with $\mathbf{r}[f,k]$.

Finally, a prefilter mask $P[f,k] \in \mathbb{C}^{E\times E}$, can be applied to emphasise particular time-frequency regimes as

$$\underline{\tilde{\mathbf{m}}}[f,k] = P[f,k]\tilde{\mathbf{m}}[f,k] \tag{10a}$$
$$= \underline{\tilde{\mathbf{s}}}[f,k] + \underline{\tilde{\mathbf{r}}}[f,k] + \underline{\tilde{\mathbf{e}}}[f,k] + \underline{\tilde{\mathbf{n}}}[f,k]. \tag{10b}$$

Similar to (10), the underline notation is used to denote prefiltered quantities, in particular, for the stacked vectors defined in (8) and (9a). When there is one desired speech source ($N_s = 1$), a common prefilter choice is the inverse of the square root of the desired speech source power spectral density (PSD), i.e., $P[f,k] = \frac{1}{\sigma_d[f,k]+\epsilon}\mathbb{I}$ with $\sigma_d[f,k] = \sqrt{\mathbb{E}\{d_1[f,k]d_1[f,k]^*\}}$, regularisation ϵ, and identity matrix $\mathbb{I}$ [5], [36], [39]. With this prefiltering, frames with a large desired speech power will contribute less to the error that will be defined on $\underline{\tilde{\mathbf{m}}}[f,k]$ in Section III. For multiple independent desired speech sources, this prefilter could be modified to $\sigma_d[f,k] = \sqrt{\sum_{n=1}^{N_s}\mathbb{E}\{d_n[f,k]d_n[f,k]^*\}}$ [36]. As per [5], [36], [39] prefilters with a diagonal structure are assumed, although non-diagonal prefilters that, e.g., leverage spatial properties, have also been derived [6].

For the above signal model, the correlation matrices $R_{yz}[f,k] = \mathbb{E}\{\mathbf{y}[f,k]\mathbf{z}[f,k]^H\}$ can be computed, for

- $\mathbf{y},\mathbf{z} \in \{\underline{\mathbf{m}}, \underline{\mathbf{m}}_D, \underline{\tilde{\mathbf{m}}}\}$ as these signals are available when $\text{VAD}[k] = 1$;
- $\mathbf{y},\mathbf{z} \in \{\underline{\mathbf{e}}+\underline{\mathbf{n}}, \underline{\mathbf{e}}_D+\underline{\mathbf{n}}_D, \underline{\tilde{\mathbf{e}}}+\underline{\tilde{\mathbf{n}}}\}$ as the echo and noise signals are available together when $\text{VAD}[k] = 0$;
- $\mathbf{y},\mathbf{z} \in \{\underline{\mathbf{s}}+\underline{\mathbf{r}}, \underline{\mathbf{s}}_D+\underline{\mathbf{r}}_D, \underline{\tilde{\mathbf{s}}}+\underline{\tilde{\mathbf{r}}}\}$ when $P[f,k] = P[f]$, i.e., under the condition that the prefilter is independent of the frame index k. The reverberant desired speech correlation matrix can then be obtained by subtracting the correlation matrix observed when $\text{VAD}[k] = 0$ from the correlation matrix observed when $\text{VAD}[k] = 1$ [37], [38];
- $\mathbf{y} \in \{\underline{\mathbf{e}}, \underline{\mathbf{e}}_D\}$ with $\mathbf{z} = \underline{\mathbf{l}}$ and $\mathbf{z} \in \{\underline{\mathbf{e}}, \underline{\mathbf{e}}_D\}$ with $\mathbf{y} = \underline{\mathbf{l}}$ as the loudspeaker signals are only correlated with the echo component in the microphone signals.

In practice, the expected value operator $\mathbb{E}\{.\}$ is replaced with averaging across frames, i.e., $R_{yz}[f,k] \approx \sum_k \mathbf{y}[f,k]\mathbf{z}[f,k]^H$ [9], [38]. Due to this averaging over a finite number of frames, the estimation of $R_{\underline{l}\underline{e}}[f,k]$ and $R_{\underline{l}\underline{e}_D}[f,k]$ is ideally inhibited when $\text{VAD}[k] = 1$ to reduce the influence of $\underline{\mathbf{s}}[f,k]$.

III. Cost Function

The goal of the integrated filter $\tilde{\mathbf{w}}_{\text{int}}[f,k] \in \mathbb{C}^E$ is to suppress the interfering echo $\tilde{\underline{\mathbf{e}}}[f,k]$, late reverberation component of the desired speech signal $\tilde{\underline{\mathbf{r}}}[f,k]$, and noise $\tilde{\underline{\mathbf{n}}}[f,k]$, while preserving the direct and early reverberation component of this desired speech signal $\tilde{\underline{\mathbf{s}}}[f,k]$. To this end, a speech distortion weighted cost function is formulated that utilises a weight μ to trade-off between echo, late reverberation and noise suppression, and preservation of the direct and early reverberation component, i.e., [11], [40]

$$\begin{aligned}\tilde{\mathbf{w}}_{\text{int}}[f,k] = \underset{\tilde{\mathbf{w}}[f,k]}{\operatorname{argmin}}\, \mathbb{E}\left\{|\underline{s}_q[f,k] - \tilde{\mathbf{w}}[f,k]^H \tilde{\underline{\mathbf{s}}}[f,k]|^2\right\} \\ + \mu \mathbb{E}\left\{|\tilde{\mathbf{w}}[f,k]^H \left(\tilde{\underline{\mathbf{r}}}[f,k] + \tilde{\underline{\mathbf{e}}}[f,k] + \tilde{\underline{\mathbf{n}}}[f,k]\right)|^2\right\},\end{aligned} \tag{11}$$

In (11), q represents the reference microphone. When $\mu = 1$, the mean squared error (MSE) is obtained [11], [40]. For $\mu > 1$, echo, late reverberation and noise suppression is emphasised, while for $\mu < 1$ preservation of the direct and early reverberation component is emphasised [11]. For conciseness, the f- and k-indices will be omitted from now on.

IV. Solution

The solution to (11) is given by the solution to the Wiener-Hopf equations

$$\underbrace{\left(R_{\tilde{\underline{s}}\tilde{\underline{s}}} + \mu\left(R_{\tilde{\underline{r}}\tilde{\underline{r}}} + R_{\tilde{\underline{e}}\tilde{\underline{e}}} + R_{\tilde{\underline{n}}\tilde{\underline{n}}}\right)\right)}_{\triangleq R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}}} \tilde{\mathbf{w}}_{\text{int}} = R_{\tilde{\underline{s}}\underline{s}}\mathbf{t}_q, \tag{12}$$

with $\mathbf{t}_q$ a unit vector with entry q equal to one. The general solution to (12) can be written as

$$\tilde{\mathbf{w}}_{\text{int}} = R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}} R_{\tilde{\underline{s}}\underline{s}}\mathbf{t}_q. \tag{13}$$

with $.^g$ denoting the generalised inverse[1,2] [41]. Indeed, as $R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ can be rank-deficient, as detailed in Section II, $R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ cannot be assumed to be invertible in general. Generalised inverses are non-unique, although specific choices can be made, such as the pseudo-inverse $.^\dagger$, which is uniquely defined [42]. When $R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ is full-rank, the generalised inverse $.^g$ can be replaced by the uniquely defined inverse $.^{-1}$.

As detailed in Section II, the direct and early reverberation component of the desired speech signal vector $\tilde{\mathbf{s}}$, cannot be observed independently from its late reverberation component $\tilde{\mathbf{r}}$. Consequently, given that (13) requires $R_{\tilde{\underline{s}}\underline{s}}$, it cannot be directly implemented. Yet, a valid generalised inverse $R^{\{\mu\}^{g,g}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ can be postulated and substituted in (13). This will result in a solution that can be implemented. By defining $\mathbf{t} \triangleq \begin{bmatrix}\mathbf{m}_D^\top & \mathbf{l}^\top\end{bmatrix}^\top$ as in (8), the 3×3 block-matrix $R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ can be rewritten as a 2×2 block matrix as

$$\begin{aligned} R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}} &= \begin{bmatrix} R^{\{\mu\}}_{\underline{m}\underline{m}} & \mu R_{\underline{m}\underline{m}_D} & \mu R_{\underline{m}\underline{l}} \\ \mu R_{\underline{m}_D\underline{m}} & \mu R_{\underline{m}_D\underline{m}_D} & \mu R_{\underline{e}_D\underline{l}} \\ \mu R_{\underline{l}\underline{e}} & \mu R_{\underline{l}\underline{e}_D} & \mu R_{\underline{l}\underline{l}} \end{bmatrix} && \text{(14a)} \\ &\triangleq \begin{bmatrix} R^{\{\mu\}}_{\underline{m}\underline{m}} & \mu R_{\underline{m}\underline{t}} \\ \mu R_{\underline{t}\underline{m}} & \mu R_{\underline{t}\underline{t}} \end{bmatrix}, && \text{(14b)} \end{aligned}$$

with $R^{\{\mu\}}_{\underline{m}\underline{m}} = R_{\underline{s}\underline{s}} + \mu\left(R_{\underline{r}\underline{r}} + R_{\underline{e}\underline{e}} + R_{\underline{n}\underline{n}}\right)$. In Supplementary Material Section VIII, it is then verified that $R^{\{\mu\}^{g,g}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ with

$$R^{\{\mu\}^{g,g}}_{\tilde{\underline{m}}\tilde{\underline{m}}} = \begin{bmatrix} \Sigma^{\{\mu\}^\dagger}_{\underline{m}\underline{m}} & -\Sigma^{\{\mu\}^\dagger}_{\underline{m}\underline{m}} R_{\underline{m}\underline{t}} R^{g,g}_{\underline{t}\underline{t}} \\ -R^{g,g}_{\underline{t}\underline{t}} R_{\underline{t}\underline{m}} \Sigma^{\{\mu\}^\dagger}_{\underline{m}\underline{m}} & \frac{1}{\mu} R^{g,g}_{\underline{t}\underline{t}} + R^{g,g}_{\underline{t}\underline{t}} R_{\underline{t}\underline{m}} \Sigma^{\{\mu\}^\dagger}_{\underline{m}\underline{m}} R_{\underline{m}\underline{t}} R^{g,g}_{\underline{t}\underline{t}} \end{bmatrix}, \tag{15}$$

is a valid generalised inverse of $R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$, where

$$R^{g,g}_{\underline{t}\underline{t}} = \begin{bmatrix} \Sigma^\dagger_{\underline{m}_D\underline{m}_D} & -\Sigma^\dagger_{\underline{m}_D\underline{m}_D} R_{\underline{e}_D\underline{l}} R^\dagger_{\underline{l}\underline{l}} \\ -R^\dagger_{\underline{l}\underline{l}} R_{\underline{l}\underline{e}_D} \Sigma^\dagger_{\underline{m}_D\underline{m}_D} & R^\dagger_{\underline{l}\underline{l}} + R^\dagger_{\underline{l}\underline{l}} R_{\underline{l}\underline{e}_D} \Sigma^\dagger_{\underline{m}_D\underline{m}_D} R_{\underline{e}_D\underline{l}} R^\dagger_{\underline{l}\underline{l}} \end{bmatrix}, \tag{16}$$

is a valid generalised inverse of $R_{\underline{t}\underline{t}}$. In (15) and (16), the following matrices are used

$$\Sigma_{\underline{m}_D\underline{m}_D} = R_{\underline{m}_D\underline{m}_D} - R_{\underline{e}_D\underline{l}} R^\dagger_{\underline{l}\underline{l}} R_{\underline{l}\underline{e}_D} \tag{17a}$$

$$\Sigma^{\{\mu\}}_{\underline{m}\underline{m}} = R^{\{\mu\}}_{\underline{m}\underline{m}} - \mu R_{\underline{m}\underline{t}} R^{g,g}_{\underline{t}\underline{t}} R_{\underline{t}\underline{m}}. \tag{17b}$$

Substituting (15) into (13) yields an expression that can be interpreted as a cascade of AEC, DR, and NR filters, i.e.,

$$\tilde{\mathbf{w}}_{\text{int}} = \underbrace{\begin{bmatrix} \mathbb{I} & 0 & 0 \\ 0 & \mathbb{I} & 0 \\ -\hat{B} & -\hat{B}_D & 0 \end{bmatrix}}_{\text{AEC}} \underbrace{\begin{bmatrix} \mathbb{I} & 0 \\ -\hat{W}_D & 0 \\ 0 & 0 \end{bmatrix}}_{\text{DR}} \underbrace{\begin{bmatrix} \hat{W}^{\{\mu\}} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{\text{NR}}. \tag{18}$$

with

$$\hat{B} = R^\dagger_{\underline{l}\underline{l}} R_{\underline{l}\underline{e}} \tag{19a}$$

$$\hat{B}_D = R^\dagger_{\underline{l}\underline{l}} R_{\underline{l}\underline{e}_D} \tag{19b}$$

$$\hat{W}_D = \Sigma^\dagger_{\underline{m}_D\underline{m}_D} \Sigma_{\underline{m}_D\underline{m}} \tag{19c}$$

$$\Sigma_{\underline{m}_D\underline{m}} = R_{\underline{m}_D\underline{m}} - R_{\underline{e}_D\underline{l}} R^\dagger_{\underline{l}\underline{l}} R_{\underline{l}\underline{e}} \tag{19d}$$

$$\hat{W}^{\{\mu\}} = \Sigma^{\{\mu\}^\dagger}_{\underline{m}\underline{m}} R_{\underline{s}\underline{s}}\mathbf{t}_q \tag{19e}$$

$$R^{g,g}_{\underline{t}\underline{t}} R_{\underline{t}\underline{m}} = \begin{bmatrix} \hat{W}_D^\top & (-\hat{B}_D\hat{W}_D + \hat{B})^\top \end{bmatrix}^\top. \tag{19f}$$

The equivalence (19f) is obtained by substituting the expression for $R^{g,g}_{\underline{t}\underline{t}}$ (16) into $R^{g,g}_{\underline{t}\underline{t}} R_{\underline{t}\underline{m}}$. The resulting filter (18), which is visualised in Fig. 1, will be shown to be implementable, as opposed to the SD-MWF$_{\text{ext}}$ in (13). Its components will be analysed in the following subsections (Section IV-A - Section IV-C), and its optimality will be further discussed in Section IV-D. Finally, Section IV-E discusses the relation between (18) and state-of-the-art solutions.

A. Acoustic Echo Cancellation (AEC)

The filter $\hat{B} \in \mathbb{C}^{LL_l \times M}$ in (19a) corresponds to the minimum-norm estimate of the echo paths $B_i(.)$, $\forall i$. Indeed, the filter $\hat{B}$ can be interpreted as the filter obtained when

[1] A generalised inverse only needs to satisfy the property $R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}} R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}} R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}} = R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ [41], thereby guaranteeing that $\tilde{\mathbf{w}}_{\text{int}}$ in (13) is a solution to the Wiener-Hopf equations (12). A pseudo-inverse needs to satisfy three additional properties: $R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}} R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}} R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}} = R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}}$, $(R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}} R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}})^H = R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}} R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}}$, and $(R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}} R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}})^H = R^{\{\mu\}^g}_{\tilde{\underline{m}}\tilde{\underline{m}}} R^{\{\mu\}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ [42].

[2] $.^g$ will be used to denote any non-specified, non-unique generalised inverse. $.^{g,g}$ will be used to denote a specific, postulated generalised inverse, according to (15) and (16).

minimising the MSE between the echo signal $\underline{\mathbf{e}}$ and the filtered loudspeaker signals $\hat{B}^H\underline{\mathbf{l}}$, i.e., $\hat{B}$ can be defined as [1]

$$\hat{B} = \underset{B}{\operatorname{argmin}}\, \mathbb{E}\left\{\left\|\underline{\mathbf{e}} - B^H\underline{\mathbf{l}}\right\|_2^2\right\}. \tag{20}$$

Similarly, $\hat{B}_D \in \mathbb{C}^{LL_l \times ML_m}$ in (19b) can be interpreted as the minimum-norm estimate, obtained when minimising the MSE between the delayed echo signal $\underline{\mathbf{e}}_D$ and the filtered loudspeaker signal $\hat{B}_D^H\underline{\mathbf{l}}$, i.e.,

$$\hat{B}_D = \underset{B_D}{\operatorname{argmin}}\, \mathbb{E}\left\{\left\|\underline{\mathbf{e}}_D - B_D^H\underline{\mathbf{l}}\right\|_2^2\right\}. \tag{21}$$

Thus, $\hat{B}$ and $\hat{B}_D$ serve as AEC filters to remove the echo from the microphone signals [1]. When $\hat{B}$ and $\hat{B}_D$ do not perfectly model the echo paths, e.g., due to non-linearities, residual echo remains.

Nevertheless, when assuming a linear echo path (4) and $L_l = L_g$, $\hat{B}$ equals $R_{\underline{ll}}^{\dagger}R_{\underline{ll}}\begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_M\end{bmatrix}$. This filter $\hat{B}$ can perfectly remove the echo, i.e., $\underline{\mathbf{e}} - \hat{B}^H\underline{\mathbf{l}} = 0$. If $R_{\underline{ll}}$ is rank-deficient, $\hat{B}$ does however not converge to the true echo path $\begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_M\end{bmatrix}$, but to a projected version thereof. This projection $R_{\underline{ll}}^{\dagger}R_{\underline{ll}}$ results in a dependence of $\hat{B}$ on the loudspeaker signals [1], [16], [43]. A similar effect is present for $\hat{B}_D$.

In practice, the updating of $R_{\underline{le}}$ and $R_{\underline{le}_D}$ is inhibited whenever the desired speech sources are active. This avoids excess error in $\hat{B}$ and $\hat{B}_D$ when the expected value operator $\mathbb{E}\{.\}$ is replaced by averaging across frames in a practical implementation (cfr. Section II-C) [1]. The prefiltering using σ_d actually fulfills this role, as the prefilter inhibits the updating of $R_{\underline{le}}$ and $R_{\underline{le}_D}$ when there is (large) desired speech source power.

B. Dereverberation (DR)

The filter $\hat{W}_D \in \mathbb{C}^{ML_m \times M}$ in (19c) is constructed using the correlation matrices $\Sigma_{\underline{m}_D\underline{m}_D}$ and $\Sigma_{\underline{m}_D\underline{m}}$. The matrix $\Sigma_{\underline{m}_D\underline{m}_D}$ can be interpreted as the auto-correlation matrix of the delayed microphone signals, following the AEC filters. Indeed, using (19b) and $R_{\underline{ll}}^{\dagger}R_{\underline{ll}}R_{\underline{ll}}^{\dagger} = R_{\underline{ll}}^{\dagger}$, as per the definition of the pseudo-inverse [41], expanding the expression within the following expected value

$$\mathbb{E}\left\{\left(\underline{\mathbf{m}}_D - \hat{B}_D^H\underline{\mathbf{l}}\right)\left(\underline{\mathbf{m}}_D - \hat{B}_D^H\underline{\mathbf{l}}\right)^H\right\} \tag{22a}$$

$$= R_{\underline{m}_D\underline{m}_D} - 2R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger}R_{\underline{ll}}R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} + R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} \tag{22b}$$

$$= R_{\underline{m}_D\underline{m}_D} - R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} = \Sigma_{\underline{m}_D\underline{m}_D}, \tag{22c}$$

results in this equivalence. Similarly, the matrix $\Sigma_{\underline{m}_D\underline{m}}$ can be interpreted as the cross-correlation matrix between the delayed microphone signals and the microphone signals, following the AEC filters. Indeed, using (19a) and (19b)

$$\mathbb{E}\left\{\left(\underline{\mathbf{m}}_D - \hat{B}_D^H\underline{\mathbf{l}}\right)\left(\underline{\mathbf{m}} - \hat{B}^H\underline{\mathbf{l}}\right)^H\right\} \tag{23a}$$

$$= R_{\underline{m}_D\underline{m}} - 2R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger}R_{\underline{ll}}R_{\underline{ll}}^{\dagger}R_{\underline{le}} + R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger}R_{\underline{le}} \tag{23b}$$

$$= R_{\underline{m}_D\underline{m}} - R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger}R_{\underline{le}} = \Sigma_{\underline{m}_D\underline{m}}. \tag{23c}$$

Consequently, $\hat{W}_D$ corresponds to a minimum-norm linear prediction filter of the microphone signals based on delayed microphone signals, following the AEC filters, i.e.,

$$\hat{W}_D = \underset{W_D}{\operatorname{argmin}}\, \mathbb{E}\left\{\left\|\left(\underline{\mathbf{m}} - \hat{B}^H\underline{\mathbf{l}}\right) - W_D^H\left(\underline{\mathbf{m}}_D - \hat{B}_D^H\underline{\mathbf{l}}\right)\right\|_2^2\right\}. \tag{24}$$

This $\hat{W}_D$ is then seen to correspond to an inverse-filtering-based multi-channel linear prediction (MCLP) filter, as widely employed for DR [5], [6], [36], [39]. When $N_s = 1$, and when the prefilter corresponds to the inverse of the square root of the desired speech source PSD, the so-called weighted prediction error (WPE) filter is obtained [5], [36], [39].

The filter $\hat{W}_D$ aims at performing DR, not only for the desired speech signal, but for all signals combined. It has been shown that in order to obtain perfect DR, $\Sigma_{\underline{m}_D\underline{m}_D}$ should be invertible [5], [36]. A necessary condition for this invertibility is [36]

$$L_m \geq \frac{(N_s + N_n + N_l)(\max(L_h, L_g, L_n) - 1)}{M - N_s - N_n - N_l}, \tag{25}$$

for N_s desired speech sources with ATFs of length L_h, N_l residual echo sources with ATFs of length L_g, and N_n noise sources with ATFs of length L_n. Additionally, the ATFs should contain no common zeros (the MINT conditions [44]), and the source signals should be temporally uncorrelated. When these conditions are not satisfied, $\hat{W}_D$ still attains DR but a bias may be present [5], [36]. When the loudspeaker signals are linearly independent, $N_l = L$. When the loudspeaker signals are linearly dependent, N_l corresponds to the number of linearly independent sources, and L_g corresponds to the length of the ATFs from this linearly independent subset of sources to the microphones. When the AEC filters perform perfect AEC, (25) can be relaxed with $N_l = 0$ and $L_g = 0$.

C. Noise Reduction (NR)

The filter $\hat{W}^{\{\mu\}} \in \mathbb{C}^M$ in (19e) is constructed using the correlation matrices $\Sigma_{\underline{mm}}^{\{\mu\}}$ and $R_{\underline{ss}}$. Similar to Section IV-B, $\Sigma_{\underline{mm}}^{\{\mu\}}$ can be interpreted as the auto-correlation matrix of the microphone signals, following the AEC and DR filters. The matrix $\Sigma_{\underline{mm}}^{\{\mu\}}$ also incorporates the weight μ from (11) for the residual echo, late reverberation, and noise. Indeed,

$$\mathbb{E}\left\{\underline{\mathbf{ss}}^H\right\} \tag{26a}$$

$$+\mu\mathbb{E}\left\{\left(\underline{\mathbf{r}}+\underline{\mathbf{e}}+\underline{\mathbf{n}}-R_{\underline{mt}}R_{\underline{tt}}^{g,g}\underline{\mathbf{t}}\right)\left(\underline{\mathbf{r}}+\underline{\mathbf{e}}+\underline{\mathbf{n}}-R_{\underline{mt}}R_{\underline{tt}}^{g,g}\underline{\mathbf{t}}\right)^H\right\}$$

$$= R_{\underline{ss}} + \mu(R_{\underline{rr}} + R_{\underline{ee}} + R_{\underline{nn}}) - 2\mu R_{\underline{mt}}R_{\underline{tt}}^{g,g}R_{\underline{tt}}R_{\underline{tt}}^{g,g}R_{\underline{tm}}$$

$$+ \mu R_{\underline{mt}}R_{\underline{tt}}^{g,g}R_{\underline{tm}} = \Sigma_{\underline{mm}}^{\{\mu\}}. \tag{26b}$$

Equation (26) relies on $R_{\underline{tt}}^{g,g}R_{\underline{tt}}R_{\underline{tt}}^{g,g} = R_{\underline{tt}}^{g,g}$, which is proven in Supplementary Material Section VIII, and on (19f). The auto-correlation matrix of the direct and early reverberation component of the desired speech signal, following the AEC and DR filters, still corresponds to $R_{\underline{ss}}$. Indeed, as in (10) the coefficients of $\tilde{\underline{\mathbf{s}}}$ related to the AEC and DR filters are zero, the AEC and DR filters do not affect $\tilde{\underline{\mathbf{s}}}$.

Consequently, $\hat{W}^{\{\mu\}}$ can be interpreted as an SD-MWF following the AEC and DR filters, i.e., [11], [40]

$$\hat{W}^{\{\mu\}} = \underset{W^{\{\mu\}}}{\operatorname{argmin}} \mathbb{E}\left\{|s_q - W^{\{\mu\}^H}\left(\underline{\mathbf{m}} - \hat{B}^H\underline{\mathbf{l}} - \hat{W}_D^H\left(\underline{\mathbf{m}}_D - \hat{B}_D^H\underline{\mathbf{l}}\right)\right)|^2\right\}. \tag{27}$$

Thus, $\hat{W}^{\{\mu\}}$ aims at reducing the noise, as well as the residual echo and residual late reverberation, while aiming to preserve the direct and early reverberation component of the desired speech signal. When the AEC and DR filters perform perfect AEC and DR, the NR filters only aim at reducing the noise.

The filter $\hat{W}^{\{\mu\}}$ is not directly implementable in general as $R_{\underline{ss}}$ is not directly accessible, given that the direct and early reverberation component of the desired speech signal is not individually observable (cfr. Section II). Yet, dependent on the choice of the weight μ and the prefilter $P[f,k]$, and dependent on the rank of the correlation matrices involved, as well as the number of desired speech sources N_s, implementable filters do exist:

1) $\mu > 0, P[f,k] = P[f]$: When the weight μ is non-zero and the prefilter is independent of the frame index, the auto-correlation matrix of the direct and early reverberation component of the desired speech signal $R_{\underline{ss}}$ can be obtained.

First, the auto-correlation matrix of the reverberant desired speech signal can be obtained as follows. When VAD$[k] = 1$, the following microphone signal vector is observed

$$\begin{aligned}\underline{\mathbf{m}} &= \underline{\mathbf{s}} + \underline{\mathbf{r}} + \underline{\mathbf{e}} + \underline{\mathbf{n}} - R_{\underline{mt}}R_{\underline{tt}}^{g,g}\underline{\mathbf{t}} \\ &= \underline{\mathbf{s}} + \underbrace{\underline{\mathbf{r}} - \hat{W}_D^H(\underline{\mathbf{s}}_D + \underline{\mathbf{r}}_D)}_{\triangleq \underline{\mathbf{r}}_{\text{res}}} + \underbrace{\underline{\mathbf{e}} - \hat{B}^H\underline{\mathbf{l}} - \hat{W}_D^H(\underline{\mathbf{e}}_D - \hat{B}_D^H\underline{\mathbf{l}})}_{\triangleq \underline{\mathbf{e}}_{\text{res}}} + \underbrace{\underline{\mathbf{n}} - \hat{W}_D\underline{\mathbf{n}}_D}_{\triangleq \underline{\mathbf{n}}_{\text{res}}},\end{aligned} \tag{28}$$

and when VAD$[k] = 0$, $\underline{\mathbf{e}}_{\text{res}} + \underline{\mathbf{n}}_{\text{res}}$ is observed. As the prefilter is independent of the frame index k, the subtraction of the correlation matrices obtained in both VAD regimes leads to $R_{\underline{ss}} + R_{\underline{r}_{\text{res}}\underline{r}_{\text{res}}}$.

Second, in [45], it has been shown that the residual late reverberation component of the desired speech signal following the DR filter can be approximated by a frame-dependent power spectral density σ_r^2 and a coherence matrix representing an isotropic sound field $\Gamma \in \mathbb{C}^{M\times M}$, i.e.,

$$R_{\underline{r}_{\text{res}}\underline{r}_{\text{res}}} = \sigma_r^2\Gamma. \tag{29}$$

This coherence matrix can be determined by assuming a diffuse coherence matrix [9], [45], by computing the coherence matrix based on the output of the DR filter [45], or by computing the coherence matrix based on measured ATFs [10].

Under this assumption of (29), $R_{\underline{ss}}$ can be extracted from the observation of $R_{\underline{ss}} + R_{\underline{r}_{\text{res}}\underline{r}_{\text{res}}}$ according to the procedure of [9]. Given (2), $R_{\underline{ss}}$ is generally of rank $R = N_sD$ for N_s independent desired speech sources that are not linearly dependent across D frames. The value of D typically resides between one and five frames [36], [39]. Consequently, when $M > R$, computing the generalised eigenvalue decomposition (GEVD) of the pencil $\{R_{\underline{ss}} + R_{\underline{r}_{\text{res}}\underline{r}_{\text{res}}}, \Gamma\}$ reveals this underlying rank-deficiency, i.e., [9]

$$R_{\underline{ss}} + R_{\underline{r}_{\text{res}}\underline{r}_{\text{res}}} = Q\Lambda Q^H \tag{30a}$$

$$= Q\left(\sigma_r^2\mathbb{I} + \text{diag}\left(\sigma_s^{(1)^2}, \cdots, \sigma_s^{(R)^2}, 0, \cdots, 0\right)\right)Q^H \tag{30b}$$

$$\Gamma = QQ^H, \tag{30c}$$

with diag(.) representing a diagonal matrix. In (30), the first R diagonal elements of $\{\sigma_r^2 + \sigma_s^{(1)^2}, \cdots, \sigma_r^2 + \sigma_s^{(R)^2}, \sigma_r^2, \cdots, \sigma_r^2\}$ are defined by both $R_{\underline{ss}}$ and $R_{\underline{r}_{\text{res}}\underline{r}_{\text{res}}}$, while σ_r^2 can be readily observed in the final $M - R$ diagonal elements [9]. By averaging these final $M - R$ diagonal elements, [9]

$$\hat{\sigma}_r^2 = \frac{1}{M-R}\left(\text{trace}(\Lambda) - \sum_{\rho=1}^{R}\left(\sigma_r^2 + \sigma_s^{(\rho)^2}\right)\right), \tag{31}$$

$R_{\underline{ss}}$ can be approximated as $\left(R_{\underline{ss}} + R_{\underline{r}_{\text{res}}\underline{r}_{\text{res}}}\right) - \hat{\sigma}_r^2\Gamma$.

2) $\mu \to 0, \Sigma_{\underline{mm}}^{\{\mu\}^\dagger} = \Sigma_{\underline{mm}}^{\{\mu\}^{-1}}, N_s = 1, D = 1$: When there is only one desired speech source ($N_s = 1$), the pseudo-inverse $\Sigma_{\underline{mm}}^{\{\mu\}^\dagger}$ corresponds to an inverse $\Sigma_{\underline{mm}}^{\{\mu\}^{-1}}$ (cfr. Section IV), $D = 1$, and μ approaches 0, the filter $\hat{W}^{\{\mu\}}$ can be implemented as an MPDR beamformer [11]. As shown in Supplementary Material Section IX, following a derivation similar to [11], [38], $\hat{W}^{\{\mu\}}$ is then indeed equivalent to the following MPDR beamformer

$$\hat{W}^{\{\mu\to 0\}} = \frac{\Sigma_{\underline{mm}}^{\{\mu=1\}^{-1}}\mathbf{h}_{\text{rel}}}{\mathbf{h}_{\text{rel}}^H\Sigma_{\underline{mm}}^{\{\mu=1\}^{-1}}\mathbf{h}_{\text{rel}}}. \tag{32}$$

In (32), $\mathbf{h}_{\text{rel}}$ collects the relative transfer functions (RTFs) $\left[\frac{h_{1,1}}{h_{1,q}} \cdots \frac{h_{1,M}}{h_{1,q}}\right]$, i.e., the ATFs of (2) normalised with respect to the ATF $h_{1,q}$ related to the reference microphone.

In (32), $\Sigma_{\underline{mm}}^{\{\mu=1\}}$ can be computed as the auto-correlation matrix of the microphone signals following the AEC and DR filters when VAD$[k] = 1$. The RTFs $\mathbf{h}_{\text{rel}}$ can be obtained according to the covariance whitening method [46], where $\mathbf{h}_{\text{rel}}$ is approximated by the normalised generalised eigenvector of the pencil $\{R_{\underline{ss}} + R_{\underline{r}_{\text{res}}\underline{r}_{\text{res}}} + R_{\underline{e}_{\text{res}}\underline{e}_{\text{res}}} + R_{\underline{n}_{\text{res}}\underline{n}_{\text{res}}}, R_{\underline{e}_{\text{res}}\underline{e}_{\text{res}}} + R_{\underline{n}_{\text{res}}\underline{n}_{\text{res}}} + \hat{\sigma}_r^2\Gamma\}$ corresponding to the largest generalised eigenvalue. As detailed in Section IV-C1, the first matrix of the pencil can be obtained when VAD$[k] = 1$. The second matrix of the pencil can be obtained by computing $R_{\underline{e}_{\text{res}}\underline{e}_{\text{res}}} + R_{\underline{n}_{\text{res}}\underline{n}_{\text{res}}}$ when VAD$[k] = 0$ and adding $\hat{\sigma}_r^2\Gamma$ of (29) to model the residual late reverberation component of the desired speech signal.

The condition $D = 1$ can in practice be relaxed by assuming $\sum_{p=0}^{D-1} h_{1,i}[f,p]^* d_n[f,k-p] \approx h_{1,i}[f,0]^* d_n[f,k]$ [18], [19], i.e., when $h_{1,i}[f,0]$ is dominant over $h_{1,i}[f,p]$, $p \in \{1,\cdots,D-1\}$. This approximation then still allows for a delay $D > 1$ to be incorporated into the DR filter, while also still allowing to implement $\hat{W}^{\{\mu\}}$ as an MPDR beamformer [19].

D. Discussion on Optimality

In Supplementary Material Section VIII, it is shown that while $R_{\tilde{m}\tilde{m}}^{\{\mu\}^{g,g}}$ of (15) is a generalised inverse, it is not generally a pseudo-inverse. Consequently, $\tilde{\mathbf{w}}_{\text{int}}$ does not necessarily correspond to the minimum-norm solution. Nevertheless, when $\Sigma_{\underline{mm}}^{\{\mu\}}$ in (17b) and $\Sigma_{\underline{m}_D\underline{m}_D}$ in (17a) are both full rank, $R_{\tilde{m}\tilde{m}}^{\{\mu\}^{g,g}}$ does correspond to a pseudo-inverse and $\tilde{\mathbf{w}}_{\text{int}}$ is the minimum-norm solution. Even when there is perfect AEC and perfect DR, this condition is satisfied when there are more than M noise sources that are not linearly related across the past L_m

frames. Indeed, due to the contribution of the noise in $\Sigma^{\{\mu\}}_{\underline{m}\underline{m}}$ and $\Sigma_{\underline{m}_D\underline{m}_D}$, both matrices are then full rank. This condition is satisfied when, e.g., (sufficiently dominant) sensor noise is present in the microphone signals [36]. When additionally, the loudspeaker correlation matrix $R_{\underline{l}\underline{l}}$ is full rank, $R^{\{\mu\}^{g,g}}_{\tilde{\underline{m}}\tilde{\underline{m}}}$ corresponds to a uniquely defined inverse.

As (18) is constructed as the cascade of AEC, DR, and NR filters, (18) can actually be rearranged in any order to arrive at, e.g., a cascade of DR, NR, and AEC filters, or a cascade of DR, AEC, and NR filters. Nevertheless, these cascades cannot be directly implemented as, e.g., the optimal DR filter of (18) makes use of (delayed) microphone correlation matrices following the AEC filters. If the DR filter precedes the AEC filters, these correlation matrices are not yet available. While these cascades can be modified to become implementable by, e.g., replacing the (delayed) echo-cancelled correlation matrices in the DR filter with the (delayed) microphone correlation matrices, they then loose their theoretical optimality. A similar result has been obtained for combined AEC and NR [16], [22].

E. Relation to State-of-the-Art Solutions

Depending on whether L_m and/or L_l are set to zero, and depending on the prefilter $P[f,k]$, equivalences between $\tilde{\mathbf{w}}_{\text{int}}$ of (18) and state-of-the-art solutions can be shown. However, note that these existing solutions rely on inverses, or at most on pseudo-inverses, rather than considering generalised inverses as in this paper, except for [16].

1) $L_l = 0$: When $L_l = 0$, the echo is treated as a regular noise source, and the information of the loudspeaker signals is not exploited, i.e., $\hat{B} = 0$ and $\hat{B}_D = 0$. Only the DR and NR filters remain. When using prefilter $P[f,k] = \frac{1}{\sigma_d+\epsilon}\mathbb{I}$ and the MPDR beamformer of (32) as NR filter, (18) becomes equivalent to the solution of [18], as originally proposed for combined DR and NR. This solution can also be equivalently implemented as a multi-frame MPDR beamformer [19].

2) $L_m = 0$: When $L_m = 0$, the delayed microphone signals are not exploited, i.e., $\hat{W}_D = 0$. The late reverberation component of the desired speech signal, instead of these signals following the DR filter, are then assumed to adhere to an isotropic sound field. Using the SD-MWF implementation for the NR filter with $\mu = 1$ and $P[f,k] = \mathbb{I}$, (18) corresponds to the cascade of AEC and NR filters of, e.g., [16], [22], as originally proposed for combined AEC and NR. Yet, with the addition that the NR filter should additionally suppress the late reverberation component of the desired speech signal. This structure of AEC and NR filters without DR filter is furthermore similar to the single-loudspeaker/single-microphone solution of [26], proposed as a cascaded solution for combined AEC, DR, and NR.

3) $L_l = 0$ and $L_m = 0$: When both $L_l = 0$ and $L_m = 0$, the loudspeaker and delayed microphone signals are not exploited, i.e., $\hat{B} = 0$, $\hat{B}_D = 0$, and $\hat{W}_D = 0$. With $P[f,k] = \mathbb{I}$, with the SD-MWF implementation for the NR filter, and with $\mu = 1$, the filter of (18) becomes equivalent to the solution of [9], as originally proposed for combined DR and NR. When selecting $P[f,k] = \frac{1}{\sigma_d+\epsilon}\mathbb{I}$, and implementing the NR filter as an MPDR beamformer, the filter of (18) becomes equivalent to the MPDR beamformer of [12], as originally proposed for NR. Finally, when selecting $P[f,k] = \mathbb{I}$, and implementing the NR filter as an MPDR beamformer, the filter of (18) becomes equivalent to the MPDR beamformer, as commonly employed for NR [38].

4) $L_l \neq 0$ and $L_m \neq 0$: In [17], the rationale of employing a cascade of AEC, DR, and NR filters has also been considered. Yet, rather than following an integrated approach, the cascade is directly postulated, and appended with a joint parameter optimisation using EM and NNs [17]. Finally, in [30], an integrated approach based on a maximum-likelihood cost function and EM has led to a similar cascade. However, this cascade consists of differently formulated correlation matrices, and with the order of the DR and AEC filters interchanged. Additionally, in [30], only a single loudspeaker and linear echo paths have been considered.

V. Experiment Design

A. Acoustic Scenarios

Thirty scenarios are generated from the MYRiAD dataset [47]. This dataset contains ATFs to both external microphones and to microphones inside a behind-the-ear (BTE) hearing instrument mounted on a dummy head [47]. A subset of $M = 5$ microphones is selected, i.e., the external microphone (XM1) as mounted on top of the dummy head, and the four BTE microphones, two at each side of the dummy head. The recording room has a reverberation time $T_{60} = 0.5\,\text{s}$. The XM1 external microphone is selected as reference microphone q.

A single desired speech source of $12\,\text{s}$ is selected, which consists of sentences of the hearing in noise test (HINT) dataset [48]. Two scenarios are considered. In Scenario I, it is a priori known that the desired speech signal is continuously active after $2\,\text{s}$, similar to [19]. In Scenario II, the HINT sentences that make up the desired speech source are interleaved with $2\,\text{s}$ of silence. Hence, a VAD is required for Scenario II, as there is no prior knowledge on the desired speech activity.

The noise source is a mixture of localised babble noise [49], and white noise added directly to the microphone signals to model sensor noise. The power ratio between the localised and the sensor noise at the reference microphone q is set to $20\,\text{dB}$. A single loudspeaker signal is used, consisting of a mixture of speech signals from the HINT dataset to model far-end room speech sources, and babble noise to model far-end room noise sources, with a power ratio of $0\,\text{dB}$ at the loudspeaker. Unless explicitly stated otherwise, the signal-to-noise ratio (SNR) and signal-to-echo ratio (SER) at the reference microphone q, defined as the power ratio between the direct and early reverberation component of the desired speech signal, and the noise and echo, respectively, are both set to $5\,\text{dB}$. The signal-to-reverberant ratio (SRR) at the reference microphone q, defined as the power ratio between the direct and early reverberation component of the desired speech signal and the late reverberation component, has a mean value of

$-1.23\,\mathrm{dB}$ with a standard deviation of $1.51\,\mathrm{dB}$ in Scenario I, and a mean value of $-0.94\,\mathrm{dB}$ with a standard deviation of $1.33\,\mathrm{dB}$ in Scenario II.

B. Algorithm Settings

The sampling frequency is set to $f_s = 16\,\mathrm{kHz}$. A square-root Hann window of $F = 256$ samples is used as STFT analysis and synthesis window. Similarly, frame overlap of $\frac{F}{2} = 128$ samples is used. The prefilter $P[f,k]$ is either omitted, i.e., $P[f,k] = \mathbb{I}$, when using the SD-MWF NR filter (Section IV-C1), or the prefilter $P[f,k] = \frac{1}{\sigma_d+\epsilon}\mathbb{I}$ with $\epsilon = 0$ is used, as per [39], when using the MPDR NR filter (Section IV-C2). For the SD-MWF NR filter, $\mu = 1$ unless explicitly stated otherwise, thus resulting in an MWF NR filter implementation.

Two processing options are explored, namely off-line and on-line processing. In off-line processing, the entire $12\,\mathrm{s}$ of audio is assumed available. The correlation matrices of (18) are then computed on the entire signal through averaging across frames. A single filter is subsequently computed, and applied to the entire signal. As per [5], [19], [39], this off-line processing allows for an iterative approach where σ_d^2 is initially estimated as the average PSD across all microphone signals in a two-sided span of five frames at each side around the current frame. After applying the filter of (18), σ_d^2 can be reestimated on the enhanced microphone signals, before recomputing the filter of (18). To this end, as per [39], σ_d^2 is estimated by averaging the PSD across all (enhanced) microphone signals, also in a two-sided span of five frames at each side around the current frame. In on-line processing, the audio is processed in a frame-by-frame basis. The correlation matrices of (18) are continuously updated, e.g.,

$$\begin{aligned} \underline{R_{mm}}[f,k] = \lambda \underline{R_{mm}}[f,k] + \\ (1-\lambda)\left((P[f,k]\mathbf{m}[f,k])(P[f,k]\mathbf{m}[f,k])^H\right), \end{aligned} \tag{33}$$

using smoothing factor λ. Upon this replacement of $\mathbb{E}\{.\}$ with on-line processing, it is assumed that the spatial scenario is sufficiently slowly varying [23], such that the same holds true for $\hat{B}$, $\hat{B}_D$, and $\hat{W}_D$, and such that (22), (23) and (26) are approximately satisfied. As per [39], σ_d^2 is estimated by averaging the PSD across all microphone signals using a span of five past frames. Unless stated otherwise, $\lambda = 0.995$ as per [50]. The auto-correlation matrices are initialised to $10^{-6}\mathbb{I}$, and the cross-correlation matrices to 0.

The filter (18) is implemented as a cascade. The correlation matrices associated with the AEC filters are continuously updated. The correlation matrices associated with the DR filter are also continuously updated, although for off-line processing, updating is disabled for the first $2\,\mathrm{s}$ as there is no desired speech activity yet. For the NR filter, in Scenario II, a VAD is employed to distinguish between the regimes of desired speech activity and inactivity. Either an oracle VAD, or the robust VAD (rVAD) of [51] is used. The subtraction of the correlation matrices, e.g., those recorded when $\mathrm{VAD}[k] = 1$ and when $\mathrm{VAD}[k] = 0$, is guided by a GEVD [37], to enforce the result to be positive semi-definite.

In off-line processing, $D = 1$ and $L_m = 4$, such that $\underline{R_{ss}}$ is theoretically a rank-1 matrix, thereby allowing the NR filter $\hat{W}^{\{\mu\}}$ to be implemented as an MPDR beamformer. For on-line processing, $D = 3$ and $L_m = 7$ to keep the same length of the DR filter $\hat{W}_D$, but to avoid distortion of the direct and early reverberation component of the desired speech signal by $\hat{W}_D$. This distortion could indeed occur due to residual correlation of $\mathbf{s}$ across frames when the assumption $\mathbb{E}\{d_1[f,k]d_1[f,k-p]\} = 0$ is not exactly satisfied [36]. Nevertheless, the rank of $\underline{R_{ss}}$ is still assumed to be one as detailed in Section IV-C2. In off-line processing $L_l = 7$, and in on-line processing $L_l = 10$, thereby exceeding L_m by three frames, in order to have a sufficient temporal span for $\hat{B}_D$.

The filter of [30], as discussed in Section IV-E, serves as a baseline. As this solution is updated using EM, iterative off-line processing is used, as per [30]. The same filter orders are used as for the derived filter of (18), the correlation matrices are similarly initialised to $10^{-6}\mathbb{I}$, and a prefilter $P[f,k] = \frac{1}{\sigma_d+\epsilon}\mathbb{I}$ is used. For the first iteration, σ_d^2 is set to one as employing the same initialisation as for the filter of (18) resulted in increased distortion of the direct and early reverberation component of the desired speech signal.

C. Performance Metrics

The intelligibility-weighted SER improvement, SRR improvement, and SNR improvement are defined, respectively, as [52], [53]

$$\Delta\mathrm{SER}^I\ [\mathrm{dB}] = \sum_{n_o=1}^{N_o} I_{n_o}\left(\mathrm{SER}_{n_o}^{\mathrm{out}} - \mathrm{SER}_{n_o}^{\mathrm{in}}\right) \tag{34a}$$

$$\Delta\mathrm{SRR}^I\ [\mathrm{dB}] = \sum_{n_o=1}^{N_o} I_{n_o}\left(\mathrm{SRR}_{n_o}^{\mathrm{out}} - \mathrm{SRR}_{n_o}^{\mathrm{in}}\right) \tag{34b}$$

$$\Delta\mathrm{SNR}^I\ [\mathrm{dB}] = \sum_{n_o=1}^{N_o} I_{n_o}\left(\mathrm{SNR}_{n_o}^{\mathrm{out}} - \mathrm{SNR}_{n_o}^{\mathrm{in}}\right), \tag{34c}$$

with

$$\mathrm{SER}_{n_o}^{\mathrm{in}} = 10\log_{10}\left(\frac{P_{s,n_o}^{\mathrm{in}}}{P_{e,n_o}^{\mathrm{in}}}\right),\ \mathrm{SER}_{n_o}^{\mathrm{out}} = 10\log_{10}\left(\frac{P_{s,n_o}^{\mathrm{out}}}{P_{e,n_o}^{\mathrm{out}}}\right), \tag{35a}$$

$$\mathrm{SRR}_{n_o}^{\mathrm{in}} = 10\log_{10}\left(\frac{P_{s,n_o}^{\mathrm{in}}}{P_{r,n_o}^{\mathrm{in}}}\right),\ \mathrm{SRR}_{n_o}^{\mathrm{out}} = 10\log_{10}\left(\frac{P_{s,n_o}^{\mathrm{out}}}{P_{r,n_o}^{\mathrm{out}}}\right), \tag{35b}$$

$$\mathrm{SNR}_{n_o}^{\mathrm{in}} = 10\log_{10}\left(\frac{P_{s,n_o}^{\mathrm{in}}}{P_{n,n_o}^{\mathrm{in}}}\right),\ \mathrm{SNR}_{n_o}^{\mathrm{out}} = 10\log_{10}\left(\frac{P_{s,n_o}^{\mathrm{out}}}{P_{n,n_o}^{\mathrm{out}}}\right), \tag{35c}$$

In (35), P_{y,n_o}^{in} represents the input power at the reference microphone q, in the n_oth one-third octave band for signal $\tilde{\mathbf{y}} \in \{\tilde{\mathbf{s}}, \tilde{\mathbf{r}}, \tilde{\mathbf{e}}, \tilde{\mathbf{n}}\}$. Similarly, P_{y,n_o}^{out} represents the output power after applying $\tilde{\mathbf{w}}_{\mathrm{int}}$. The intelligibility weighting refers to weighting the SER, SRR, and SNR in each one-third octave band using band importance I_{n_o} as defined in [54, Table 3] with $N_o = 18$. A higher $\Delta\mathrm{SER}^I$, $\Delta\mathrm{SRR}^I$, and $\Delta\mathrm{SNR}^I$ indicates a higher performance.

To further characterise the DR performance, the cepstral distance (CD) improvement, $\Delta\mathrm{CD}$, is used [55]. As reference signal for the CD, the direct and early reverberation component of the desired speech signal in reference microphone q ($\tilde{\mathbf{t}}_q^\top\tilde{\mathbf{s}}$) is used. For the input CD, this reference signal is compared to the reverberant desired speech signal in reference microphone

q ($\tilde{\mathbf{t}}_q^\top(\tilde{\mathbf{s}}+\tilde{\mathbf{r}})$). For the output CD, this reference signal is compared to the filtered version of the reverberant desired speech signal ($\tilde{\mathbf{w}}_{\text{int}}^H(\tilde{\mathbf{s}}+\tilde{\mathbf{r}})$). To calculate the CD, the implementation of the REVERB challenge is used [55]. A lower ΔCD indicates a higher performance.

To characterise the preservation of the direct and early reverberation component of the desired speech signal, the intelligibility-weighted speech distortion (SD),

$$\text{SD}^I\ [\text{dB}] = \sum_{n_o=1}^{N_o} I_{n_o} 10\log_{10}\left(\frac{P_{s,n_o}^{\text{in}}}{P_{s,n_o}^{\text{out}}}\right), \tag{36}$$

is used [52], [53]. As $\tilde{\mathbf{s}}$ has a 0-structure corresponding to the delayed microphone and loudspeaker signals, SD^I characterises the distortion introduced by the NR filter. An SD^I closer to zero indicates a higher performance.

To characterise the aggregated effect on speech quality, the wideband perceptual evaluation of speech quality (PESQ) improvement, ΔPESQ, is used [56]. To characterise the aggregated effect on speech intelligibility, the extended short-time objective intelligibility (ESTOI) improvement, ΔESTOI, is used [57]. As reference signal, as with the ΔCD, the direct and early reverberation component of the desired speech signal in reference microphone q ($\tilde{\mathbf{t}}_q^\top\tilde{\mathbf{s}}$) is used. For the input PESQ and ESTOI, this reference signal is compared to the microphone signal in reference microphone q ($\tilde{\mathbf{t}}_q^\top\tilde{\mathbf{m}}$). For the output PESQ and ESTOI, this reference signal is compared to the filtered version of the microphone signals ($\tilde{\mathbf{w}}_{\text{int}}^H\tilde{\mathbf{m}}$). A higher ΔPESQ and a higher ΔESTOI indicates a higher performance.

VI. Results and Discussion

A. Comparison and Influence of Iterations

Fig. 2 depicts the results using off-line processing for Scenario I using the derived filter (18) with the MPDR-based NR filter across multiple iterations. Fig. 2 also depicts the results of the baseline filter of [30]. The derived filter effectively suppresses the echo, late reverberation, and noise signals given the positive ΔSER^I, ΔSRR^I and ΔSNR^I, and given the negative ΔCD. Further, given the SD^I close to zero, the direct and early reverberation component of the desired speech signal is not altered substantially. These results are fused into a positive ΔPESQ and ΔESTOI.

Additionally, the performance of the derived filter remains essentially constant across the iterations, with only a slight increase in mean ΔPESQ of 0.02 and in mean ΔESTOI of 0.02 from the first to the second iteration. Consequently, the prefilter as computed on the microphone signals reaches effective performance, facilitating on-line processing.

Finally, the derived filter (18) outperforms the baseline of [30], reaching increased ΔSER^I, ΔSRR^I and ΔSNR^I, while reaching a comparable SD^I. This performance improvement is also reflected in an increased ΔPESQ and ΔESTOI compared to the baseline of [30]. The DR performance decrease of the baseline of [30], as indicated by a positive ΔCD, seems to be caused by the sensitivity of the baseline to the choice of D and L_m. Indeed, the ΔCD becomes negative for larger values of both parameters, e.g., a mean ΔCD of -0.37 and a mean ΔESTOI of 0.14 are achieved when increasing D to 2 and L_m to 8.

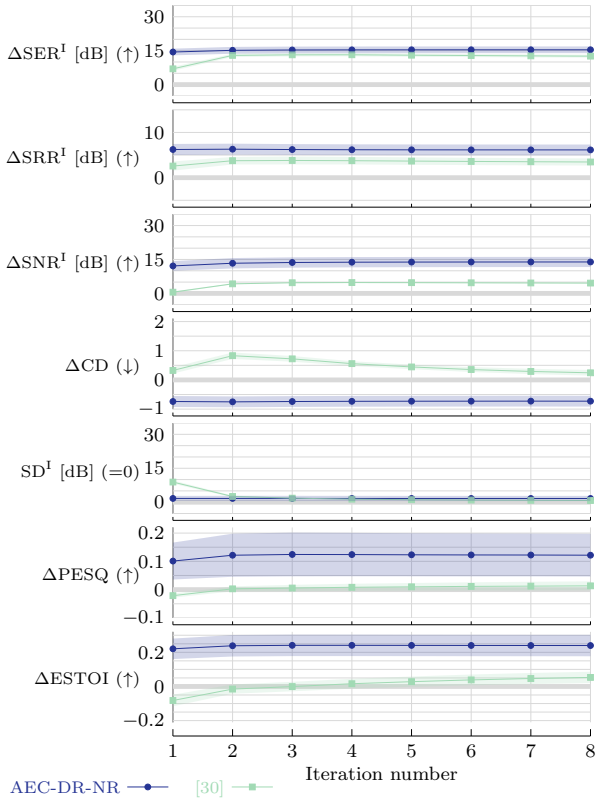

Fig. 2. Comparison between the off-line processing of the derived solution (AEC-DR-NR) with the MPDR-based NR filter, and the filter of [30] ([30]) for Scenario I. The mean value is represented with a line with markers, and the standard deviation is represented with shading. The symbols (↑), (↓), and (= 0) indicate that higher values, lower values, or values closer to zero are desired.

B. Hyperparameter Influence

Fig. 3 and Fig. 4 depict the results when using on-line processing for Scenario I. In Fig. 3, an MPDR NR filter is employed. In Fig. 4, an (SD-)MWF (μ = 1) NR filter is employed. As detailed in Section IV-C and in Section V-B, a prefilter is employed when using the MPDR NR filter, while no prefilter is employed when using the MWF NR filter. The delayed microphone and loudspeaker signals are either included or excluded by setting L_m and/or L_l to zero. The time constant τ that is coupled to the smoothing factor λ is varied [50]. This leads to the following effect of the hyperparameters.

1) Prefiltering: Each combination in Fig. 3 and Fig. 4 reaches positive ΔSNR^I and ΔSER^I at large enough time constants, indicating that both the AEC and NR filters are effective with or without the prefilter. However, the ΔCD becomes positive when employing a DR filter while omitting the prefiltering, even at large time constants. This effect can be seen in Fig. 4 when comparing the cascade of AEC, DR and NR filters to the cascade of AEC and NR filters, or the cascade of DR and NR filters to the NR filter. Indeed, as analysed in [5], [20], [36], for the DR filter to be effective, a proper prefilter should be employed, especially for STFT-domain implementations. Although the MWF NR filter is

also aimed at reducing reverberation, and hence is aimed at reducing the effect of an improper DR filter, performance is increased when leaving out the DR filter without prefiltering.

Consequently, while the AEC and NR filters are effective, even without prefiltering, the prefiltering is crucial for the DR filter. It is still possible to employ a DR filter that includes prefiltering with an MWF NR filter that does not include prefiltering, although this modified filter is then not exactly equivalent to (13).

2) Smoothing Factor: When solely using the NR filter for DR, without prior DR filter, typically small time constants are used to track the fast-varying temporal characteristics of the late reverberation component of the desired speech signal [9]. However, the AEC and DR filters, and NR filters for NR typically employ larger time constants to improve the (spatial) estimates of the correlation matrices under the assumption that the (spatial) characteristics of the scenario are slowly varying [23], [50]. These effects are reflected in Fig. 3 and Fig. 4. The solutions that employ the DR filter, which models the inverse of the spatial ATFs [5], [6], [36], display the largest performance drop at small time constants. Nevertheless, for all filters, the ΔSER^I and ΔSNR^I performance drops at small time constants due to a reduced number of data points used to compute the correlation matrices. When solely employing the NR filter, the DR performance seems less affected by the time constant, as indicated by the ΔSRR^I and ΔCD, compared to the AEC and NR performance, as indicated by the ΔSER^I and ΔSNR^I.

3) L_m and L_l: When omitting the loudspeaker signals, i.e., setting L_l to zero, the AEC performance decreases compared to utilising a non-zero L_l, certainly at large time constants, as indicated by the ΔSER^I in Fig. 3 and Fig. 4. Thus, using the AEC filters improves the AEC performance. Similarly, when omitting the delayed microphone signals, i.e., setting L_m to zero, the DR performance decreases compared to utilising a non-zero L_m at large time constants and when using the prefilter, indicated by the ΔSRR^I and ΔCD in Fig. 3.

4) Speech Distortion Weight: Fig. 5 depicts the results for the on-line processing for Scenario I using the SD-MWF NR filter, and hence without prefilter, as a function of the weight μ. As per Section VI-B1, the DR filter has been omitted given the performance reduction when employing the DR filter without prefiltering. Consistent with the results for the SD-MWF for plain NR [11], [40], increasing μ allows to trade-off between increasing ΔSER^I, ΔSRR^I and ΔSNR^I versus reducing SD^I.

C. Signal-to-Noise Ratio (SNR) and Signal-to-Echo Ratio (SER)

Fig. 6 depicts the results for on-line processing with an MPDR NR filter, and hence with prefilter, for Scenario I, as a function of the input SER. The input SER ranges from $-20\,\text{dB}$ to $20\,\text{dB}$ with a step size of $10\,\text{dB}$. The input SNR is changed from $5\,\text{dB}$ to $0\,\text{dB}$, such that the noise dominates over the echo for positive input SERs, and such that the echo dominates

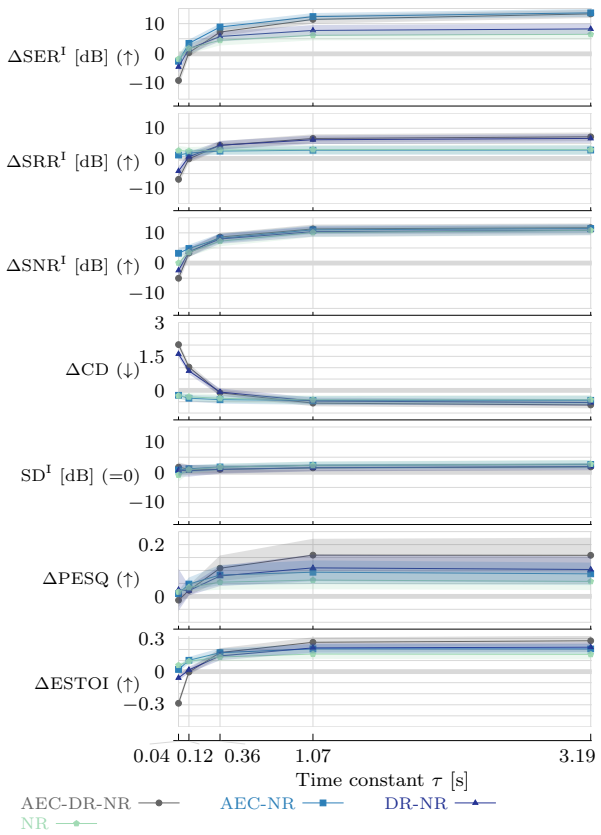

Fig. 3. Performance using on-line processing (AEC-DR-NR) for Scenario I with the MPDR NR filter, and with prefilter. When $L_m = 0$, the DR filter is omitted (AEC-NR), and when $L_l = 0$, the AEC filters are omitted (DR-NR). When both orders are set to zero, the AEC and DR filters are omitted (NR).

over the noise for negative input SERs. Next to the filter (18), alternate arrangements (Section IV-D) of the AEC, DR, and NR filters are included.

For the low input SER scenarios, the derived filter of (18) shows improved performance with respect to the alternate arrangements. As the AEC filters are well converging, due to the echo dominating the other signals, they already partially suppress the dominant echo, such that the NR filter and the DR filter can focus on reducing the noise and late reverberation component of the desired speech signal, leading to an improvement in ΔSNR^I and ΔSRR^I, and to an overall improvement in ΔESTOI. At high input SER, the three different cascaded implementations show little difference.

For all arrangements, at low input SER, the echo dominates over the noise. Consequently, the AEC filters are well-convergent, and the NR filter mainly aims at reducing residual echo, leading to a higher ΔSER^I than ΔSNR^I. The reverse trend is observed at high input SER.

D. Voice Activity Detection (VAD)

Fig. 7 depicts the results for on-line processing with an MPDR NR filter, and hence with prefilter, for Scenario II. As Scenario II is evaluated, a VAD is required to detect the activity of the desired speech source. More specifically, the VAD is required to compute $\Sigma_{\underline{mm}}^{\{\mu=1\}}$ as in (32), and to apply the

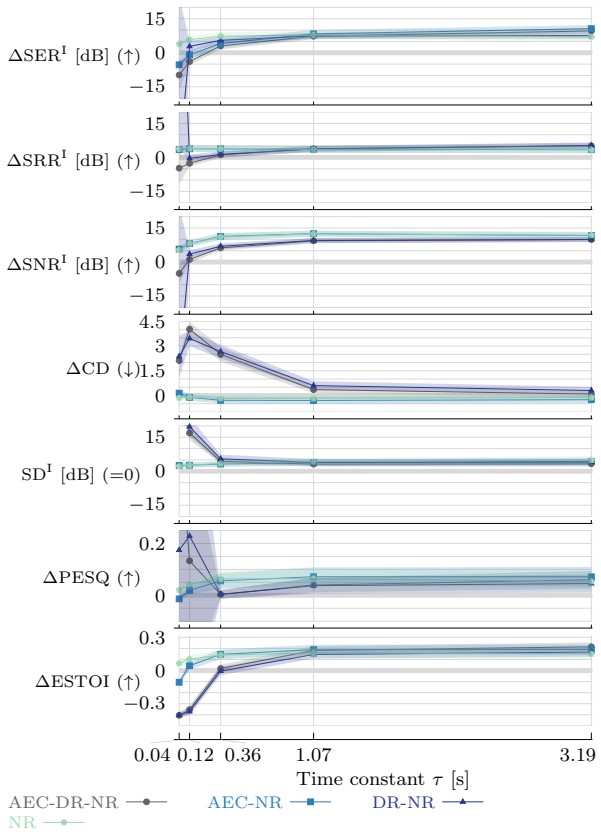

Fig. 4. Performance using on-line processing (AEC-DR-NR) for Scenario I with the MWF NR filter, and without prefilter. When $L_m = 0$, the DR filter is omitted (AEC-NR), and when $L_l = 0$, the AEC filters are omitted (DR-NR). When both orders are set to zero, the AEC and DR filter are omitted (NR).

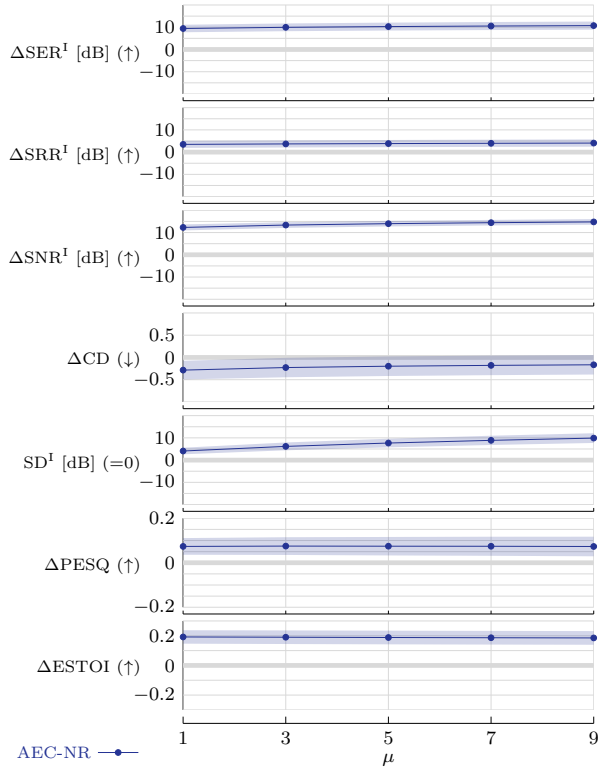

Fig. 5. Performance using on-line processing for Scenario I when using the SD-MWF NR filter without prefilter, and when omitting the DR filter, as a function of the weight μ.

the covariance whitening method to estimate $\mathbf{h}_{\text{rel}}$ as detailed in Section IV-C2 [46]. In Fig. 7, results are depicted using both an oracle VAD and a practical, imperfect VAD, the rVAD [51]. The performance when using the rVAD only slightly reduces with respect to using the oracle VAD, such that the derived filter demonstrates a degree of robustness against VAD errors.

VII. Conclusion

In this paper, an integrated algorithm design approach has been proposed for the problem of combined acoustic echo cancellation (AEC), dereverberation (DR), and noise reduction (NR). A general multi-microphone/multi-loudspeaker scenario with possible linear dependencies between (delayed) microphone and loudspeaker signals was considered. In this integrated approach, a single signal model is formulated, together with a common cost function. A common solution strategy leads to a speech distortion-weighted multi-channel Wiener filter on the extended signal model, referred to as SD-MWF$_{\text{ext}}$. This SD-MWF$_{\text{ext}}$ is shown to be equivalent to a cascade of AEC, DR, and NR filters.

The derived solution has been shown to effectively suppress the echo, late reverberation and noise, while retaining the direct component and early reverberation of the desired speech signal, both using off-line and on-line processing. A prefilter, as commonly employed in DR algorithms, improves the performance, especially for the DR filter. For on-line processing, the derived solution showed better performance at large time constants. Finally, it has been shown to be robust against voice activity detector (VAD) errors.

References

[1] E. Hänsler and G. Schmidt, *Topics in acoustic echo and noise control*, 1st ed. Berlin, Heidelberg, Germany: Springer Berlin, Heidelberg, May 2006.

[2] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, 1st ed., ser. Signals and Communication Technology. London, UK: Springer London, Jul. 2010.

[3] E. Seidel, G. Enzner, P. Mowlaee, and T. Fingscheidt, "Neural Kalman filters for acoustic echo cancellation: Comparison of deep neural network-based extensions [Special issue on model-based and data-driven audio signal processing]," *IEEE Signal Process. Mag.*, vol. 41, no. 6, pp. 24–38, Nov. 2024.

[4] K. Xie, Z. Yang, J. Chen, and J. Li, "Attention-based dual stream interactive network for nonlinear residual echo suppression," in *Proc. 2024 32nd Eur. Signal Process. Conf. (EUSIPCO)*, Lyon, France, Aug. 2024, pp. 221–225.

[5] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and Biing-Hwang Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 7, pp. 1717–1731, Sep. 2010.

[6] T. Yoshioka and T. Nakatani, "Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 10, pp. 2707–2720, Dec. 2012.

[7] J. Heymann, L. Drude, R. Haeb-Umbach, K. Kinoshita, and T. Nakatani, "Joint optimization of neural network-based WPE dereverberation and acoustic model for robust online ASR," in *Proc. 2021 IEEE Int. Conf. Acoust. Speech, Signal Proces. (ICASSP)*, Brighton, UK, May 2019, pp. 6655–6659.

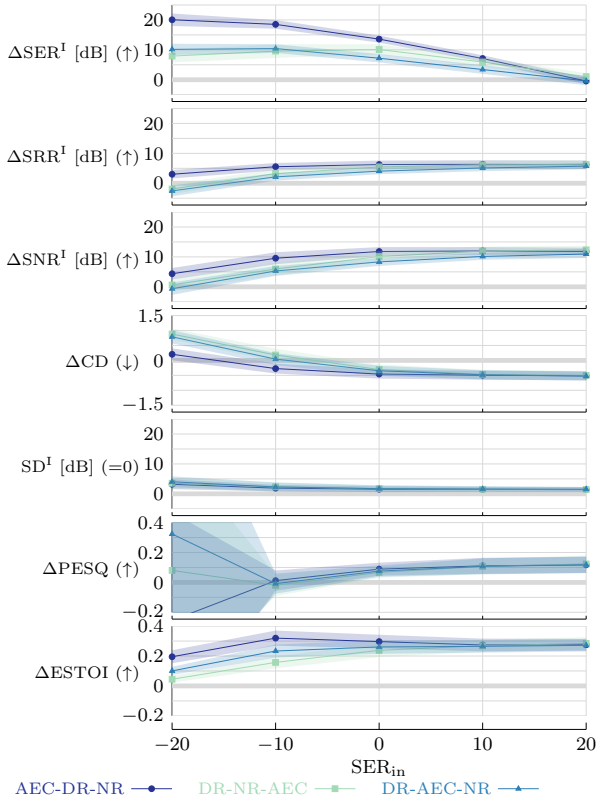

Fig. 6. Performance of the derived filter (AEC-DR-NR) using on-line processing, MPDR NR filter, and prefilter for Scenario I as a function of the input SER. The cascaded implementations of DR, NR, and AEC filters (DR-NR-AEC), and of DR, AEC, and NR filters (DR-AEC-NR) are shown as well.

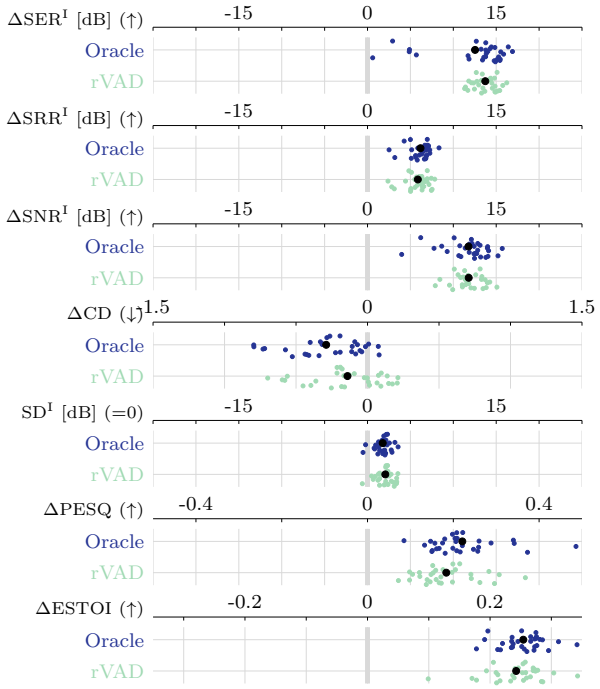

Fig. 7. Performance of the derived filter when using on-line processing for Scenario II with MPDR NR filter and prefilter. Either an oracle VAD (oracle) or practical VAD (rVAD) is used. Coloured dots represent individual scenarios, and black dots the mean value across scenario.

[8] Z.-Q. Wang, G. Wichern, and J. Le Roux, "Convolutive prediction for monaural speech dereverberation and noisy-reverberant speaker separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 29, pp. 3476–3490, Nov. 2021.

[9] I. Kodrasi and S. Doclo, "Analysis of eigenvalue decomposition-based late reverberation power spectral density estimation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 6, pp. 1106–1118, Jun. 2018.

[10] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids," in *Proc. 2014 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Lisbon, Portugal, Sep. 2014, pp. 61–65.

[11] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.

[12] B. J. Cho, J.-M. Lee, and H.-M. Park, "A beamforming algorithm based on maximum likelihood of a complex Gaussian distribution with time-varying variances for robust speech recognition," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1398–1402, Sep. 2019.

[13] Z. Zhang, Y. Xu, M. Yu, S.-X. Zhang, L. Chen, and D. Yu, "ADL-MVDR: All deep learning MVDR beamformer for target speech separation," in *Proc. 2021 IEEE Int. Conf. Acoust. Speech Signal Proces. (ICASSP)*, Toronto, ON, Canada, Jun. 2021, pp. 6089–6093.

[14] W. Jiang, C. Sun, F. Chen, Y. Leng, and Q. Guo, "A novel skip connection mechanism based on channel-wise cross transformer for speech enhancement," *Multimed. Tools appl.*, vol. 83, no. 12, pp. 34 849–34 866, Sep. 2023.

[15] H. Wang, A. Pandey, and D. Wang, "A systematic study of DNN based speech enhancement in reverberant and reverberant-noisy environments," *Comput. Speech Lang.*, vol. 89, p. 101677, Jun. 2024.

[16] A. Roebben, T. van Waterschoot, J. Wouters, and M. Moonen, "Integrated minimum mean squared error algorithms for combined acoustic echo cancellation and noise reduction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 34, pp. 512–528, Dec. 2025.

[17] G. Carbajal, R. Serizel, E. Vincent, and E. Humbert, "Joint NN-supported multichannel reduction of acoustic echo, reverberation and noise," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 2158–2173, Jul. 2020.

[18] T. Nakatani and K. Kinoshita, "A unified convolutional beamformer for simultaneous denoising and dereverberation," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 903–907, Jun. 2019.

[19] C. Boeddeker, T. Nakatani, K. Kinoshita, and R. Haeb-Umbach, "Jointly optimal dereverberation and beamforming," in *Proc. 2020 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 216–220.

[20] T. Dietzen, S. Doclo, M. Moonen, and T. van Waterschoot, "Integrated sidelobe cancellation and linear prediction Kalman filter for joint multi-microphone speech dereverberation, interfering speech cancellation, and noise reduction," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 740–754, Jan. 2020.

[21] W. Herbordt, W. Kellermann, and S. Nakamura, "Joint optimization of LCMV beamforming and acoustic echo cancellation," in *Proc. 12th Eur. Signal Process. Conf. (EUSIPCO)*, Vienna, Austria, Sep. 2004, pp. 2003–2006.

[22] S. Gustafsson, R. Martin, and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Process.*, vol. 64, no. 1, pp. 21–32, Jan. 1998.

[23] G. Rombouts and M. Moonen, "An integrated approach to acoustic noise and echo cancellation," *Signal Process.*, vol. 85, no. 4, pp. 849–871, Apr. 2005.

[24] K. Nathwani, "Joint acoustic echo and noise cancellation using spectral domain Kalman filtering in double-talk scenario," in *Proc. 16th Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Tokyo, Japan, Sep. 2018, pp. 1–330.

[25] R. Martin and P. Vary, "Combined acoustic echo cancellation, dereverberation and noise reduction: A two microphone approach," *Ann. Télécommunications*, vol. 49, pp. 429–438, Jul. 1994.

[26] E. A. P. Habets, S. Gannot, I. Cohen, and P. C. W. Sommen, "Joint dereverberation and residual echo suppression of speech signals in noisy environments," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 8, pp. 1433–1451, Nov. 2008.

[27] S. Xu, C. He, B. Yan, and M. Wang, "A multi-stage acoustic echo cancellation model based on adaptive filters and deep neural networks," *Electronics*, vol. 12, no. 15, p. 3258, Jan. 2023.

[28] X. Sun, C. Cao, Q. Li, L. Wang, and F. Xiang, "Explore relative and

context information with transformer for joint acoustic echo cancellation and speech enhancement," in *Proc. 2022 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Singapore, Singapore, May 2022, pp. 9117–9121.

[29] N. Cohen, G. Hazan, B. Schwartz, and S. Gannot, "An online algorithm for echo cancellation, dereverberation and noise reduction based on a Kalman-EM method," *EURASIP J. Audio Speech Music Process.*, vol. 2021, no. 1, p. 33, Aug. 2021.

[30] M. Togami and Y. Kawaguchi, "Simultaneous optimization of acoustic echo reduction, speech dereverberation, and noise reduction against mutual interference," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 11, pp. 1612–1623, Nov. 2014.

[31] M. Togami, "Variational Bayes state space model for acoustic echo reduction and dereverberation," in *Proc. 2015 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, South Brisbane, Queensland, Australia, Apr. 2015, pp. 101–105.

[32] N. C. Ristea, E. Indenbom, A. Saabas, T. Pärnamaa, J. Guzhvin, and R. Cutler, "DeepVQE: Real time deep voice quality enhancement for joint acoustic echo cancellation, noise suppression and dereverberation," Dublin, Ireland, pp. 3819–3823, Aug. 2023.

[33] Y. Jiang, B. Tian, H. Wang, S. Zhao, B. Ma, D. Chen, and X. Li, "E2E-AEC: Implementing an end-to-end neural network learning approach for acoustic echo cancellation," in *Proc. 2026 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2026.

[34] N. Fleischhacker, "The tikzpeople package," 2017. [Online]. Available: https://ctan.org/pkg/tikzpeople

[35] A. Roebben, "Github repository: Integrated minimum mean squared error algorithms for combined acoustic echo cancellation, dereverberation, and noise reduction,," https://cloud.esat.kuleuven.be/index.php/s/dsJxfj4W3BTSoE7, 2026.

[36] T. Dietzen, A. Spriet, W. Tirry, S. Doclo, M. Moonen, and T. van Waterschoot, "Comparative analysis of generalized sidelobe cancellation and multi-channel linear prediction for speech dereverberation and noise reduction," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 3, pp. 544–558, Mar. 2019.

[37] 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 Trans. Audio Speech Lang. Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014.

[38] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*. John Wiley & Sons, Ltd, Jan. 2010, ch. 9, pp. 269–302.

[39] L. Drude, J. Heymann, C. Boeddeker, and R. Haeb-Umbach, "NARA-WPE: A Python package for weighted prediction error dereverberation in Numpy and Tensorflow for online and offline processing," in *Speech Comm.; 13th ITG-Symposium*, Oldenburg, Germany, Oct. 2018, pp. 1–5.

[40] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction," in *Speech Enhancement*. Berlin/Heidelberg: Springer-Verlag, Mar. 2005, pp. 199–228.

[41] M. James, "The generalised inverse," *Math. Gaz.*, vol. 62, no. 420, pp. 109–114, Jun. 1978.

[42] R. Penrose, "A generalized inverse for matrices," *Math. Proc. Cambridge Philos. Soc.*, vol. 51, no. 3, pp. 406–413, Jul. 1955.

[43] S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," in *Proc. 1995 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 5, Detroit, MI, USA, May 1995, pp. 3059–3062.

[44] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[45] A. Cohen, G. Stemmer, S. Ingalsuo, and S. Markovich-Golan, "Combined weighted prediction error and minimum variance distortionless response for dereverberation," in *Proc. 2017 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 446–450.

[46] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *Proc. 2015 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, South Brisbane, Queensland, Australia, Apr. 2015, pp. 544–548.

[47] T. Dietzen, R. Ali, M. Taseska, and T. van Waterschoot, "MYRiAD: A multi-array room acoustic database," *EURASIP J. Audio Speech Music Process.*, vol. 2023, no. 17, p. 17, Apr. 2023.

[48] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *Acoust. Soc. Amer.*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

[49] Auditec, "Auditory tests (revised), compact disc, auditec," St. Louis, MO, 1997.

[50] R. Ali, "Multi-microphone speech enhancement: An integration of a priori and data-dependent spatial information," Ph.D. dissertation, KU Leuven, Leuven, Belgium, 2020.

[51] Z.-H. Tan, A. kr. Sarkar, and N. Dehak, "rVAD: An unsupervised segment-based robust voice activity detection method," *Comput. Speech Lang.*, vol. 59, pp. 1–21, Jan. 2020.

[52] A. Spriet, "Adaptive filtering techniques for noise reduction and acoustic feedback cancellation in hearing aids," Ph.D. dissertation, KU Leuven, Leuven, Belgium, 2004.

[53] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *Acoust. Soc. Amer.*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[54] Acoustical Society of America, "ANSI S3.5-1997 American national standard methods for calculation of the speech intelligibility index," ANSI S3.5-1997, Washington D.C., Washington, USA, 1997.

[55] K. Kinoshita, M. Delcroix, T. Yoshioka, T. Nakatani, E. Habets, R. Haeb-Umbach, V. Leutnant, A. Sehr, W. Kellermann, R. Maas, S. Gannot, and B. Raj, "The reverb challenge: A common evaluation framework for dereverberation and recognition of reverberant speech," in *Proc. 2013 IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 2013, pp. 1–4.

[56] T. S. S. I.-T. International Telecommunication Union, "Wideband extension to recommendation p.862 for the assessment of wideband telephone networks and speech codecs," P.862.2, Geneva, Switzerland, 2007.

[57] J. Jensen and C. H. Taal, "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 11, pp. 2009–2022, Nov. 2016.

Integrated Minimum Mean Squared Error Algorithms for Combined Acoustic Echo Cancellation, Dereverberation, and Noise Reduction: Supplementary Material

Arnout Roebben, *Graduate Student Member, IEEE*, Toon van Waterschoot, *Member, IEEE*, Jan Wouters, and Marc Moonen, *Fellow, IEEE*

For consistency with the main text, all equation, citation, and section numbers in this supplementary material are matching with those in the main text.

In Section VIII, it will be shown that (15) corresponds to a valid generalised inverse in general, and to a pseudo-inverse when $\text{rank}(\Sigma_{\underline{mm}}^{\{\mu\}}) = M$ and $\text{rank}(\Sigma_{\underline{m}_D\underline{m}_D}) = ML_m$. In Section IX, it will be shown that $\hat{W}^{\{\mu\}}$ of (19e) corresponds to the minimum power distortionless response (MPDR) beamformer of (32) when there is only one desired speech source ($N_s = 1$), $\Sigma_{\underline{mm}}^{\{\mu\}^\dagger}$ corresponds to an inverse $\Sigma_{\underline{mm}}^{\{\mu\}^{-1}}$, $D = 1$, and μ approaches zero.

VIII. Proof of (15)

In this section, it will be shown that $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$ of (15) is a valid generalised inverse of $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}$, i.e., that the necessary and sufficient condition $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} = R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}$ is always satisfied [41]. In addition, it will be shown that $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$ of (15) is a pseudo-inverse of $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}$ when $\text{rank}(\Sigma_{\underline{mm}}^{\{\mu\}}) = M$ and $\text{rank}(\Sigma_{\underline{m}_D\underline{m}_D}) = ML_m$, i.e., that the additional conditions $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} = R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$, $(R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}})^H = R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$, and $(R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}})^H = R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}$ are satisfied under these assumptions [42].

The proofs make use of the following decomposition of $R_{\underline{tt}}$,

$$R_{\underline{tt}} = \begin{bmatrix} \mathbb{I} & R_{\underline{e}_D\underline{l}} R_{\underline{ll}}^\dagger \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D} & 0 \\ 0 & R_{\underline{ll}} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{ll}}^\dagger R_{\underline{le}_D} & \mathbb{I} \end{bmatrix}, \tag{36}$$

This decomposition relies on the property $R_{\underline{ll}}^\dagger R_{\underline{ll}} R_{\underline{ll}}^\dagger = R_{\underline{ll}}^\dagger$ as per the definition of the pseudo-inverse [42]. Additionally, it relies on the property $R_{\underline{el}} R_{\underline{ll}}^\dagger R_{\underline{ll}} = R_{\underline{el}}$, as the row space of $R_{\underline{el}}$ is contained in the row space of $R_{\underline{ll}}$ and as $R_{\underline{ll}}^\dagger R_{\underline{ll}}$ is a projection matrix on the row space of $R_{\underline{ll}}$ [59, Chapter 5]. The postulated generalised inverse of $R_{\underline{tt}}$, i.e., $R_{\underline{tt}}^{g,g}$ (16), can similarly be decomposed as

$$R_{\underline{tt}}^{g,g} = \begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{ll}}^\dagger R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D}^\dagger & 0 \\ 0 & R_{\underline{ll}}^\dagger \end{bmatrix} \begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D\underline{l}} R_{\underline{ll}}^\dagger \\ 0 & \mathbb{I} \end{bmatrix}, \tag{37}$$

The matrix $R_{\underline{tt}}^{g,g}$ of (37) is indeed a generalised inverse. Using (36) and (37), $R_{\underline{tt}} R_{\underline{tt}}^{g,g} R_{\underline{tt}}$ can be expanded as

$$\begin{aligned} R_{\underline{tt}} R_{\underline{tt}}^{g,g} R_{\underline{tt}} = & \begin{bmatrix} \mathbb{I} & R_{\underline{e}_D\underline{l}} R_{\underline{ll}}^\dagger \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D} & 0 \\ 0 & R_{\underline{ll}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{ll}}^\dagger R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{ll}}^\dagger R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D}^\dagger & 0 \\ 0 & R_{\underline{ll}}^\dagger \end{bmatrix} \begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D\underline{l}} R_{\underline{ll}}^\dagger \\ 0 & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & R_{\underline{e}_D\underline{l}} R_{\underline{ll}}^\dagger \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D} & 0 \\ 0 & R_{\underline{ll}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{ll}}^\dagger R_{\underline{le}_D} & \mathbb{I} \end{bmatrix}, \end{aligned} \tag{38}$$

and can be simplified using

$$\begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{ll}}^\dagger R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{ll}}^\dagger R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} = \mathbb{I} \tag{39a}$$

$$\begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D\underline{l}} R_{\underline{ll}}^\dagger \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & R_{\underline{e}_D\underline{l}} R_{\underline{ll}}^\dagger \\ 0 & \mathbb{I} \end{bmatrix} = \mathbb{I}, \tag{39b}$$

to

$$\begin{aligned} R_{\underline{tt}} R_{\underline{tt}}^{g,g} R_{\underline{tt}} = & \begin{bmatrix} \mathbb{I} & R_{\underline{e}_D\underline{l}} R_{\underline{ll}}^\dagger \\ 0 & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D} \Sigma_{\underline{m}_D\underline{m}_D}^\dagger \Sigma_{\underline{m}_D\underline{m}_D} & 0 \\ 0 & R_{\underline{ll}} R_{\underline{ll}}^\dagger R_{\underline{ll}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{ll}}^\dagger R_{\underline{le}_D} & \mathbb{I} \end{bmatrix}. \end{aligned} \tag{40}$$

As per the definition of the pseudo-inverse, $\Sigma_{\underline{m}_D\underline{m}_D} \Sigma_{\underline{m}_D\underline{m}_D}^\dagger \Sigma_{\underline{m}_D\underline{m}_D} = \Sigma_{\underline{m}_D\underline{m}_D}$ and $R_{\underline{ll}} R_{\underline{ll}}^\dagger R_{\underline{ll}} = R_{\underline{ll}}$. Consequently, $R_{\underline{tt}} R_{\underline{tt}}^{g,g} R_{\underline{tt}} = R_{\underline{tt}}$.

This research was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven project C14-21-0075 'A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices', and Aspirant Grant 11PDH24N (for A. Roebben) from the Research Foundation - Flanders (FWO). The scientific responsibility is assumed by its authors.

Arnout Roebben, Toon van Waterschoot and Marc Moonen are with the Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, B-3001 Leuven, Belgium (e-mail: arnout.roebben@esat.kuleuven.be; toon.vanwaterschoot@esat.kuleuven.be; marc.moonen@esat.kuleuven.be).

Jan Wouters is with the Department of Neurosciences, Research Group ExpORL, KU Leuven, B-3000 Leuven, Belgium (e-mail: jan.wouters@med.kuleuven.be).

The generalised inverse $R^{g,g}_{\underline{tt}}$ additionally satisfies the property $R^{g,g}_{\underline{tt}} R_{\underline{tt}} R^{g,g}_{\underline{tt}} = R^{g,g}_{\underline{tt}}$. Indeed,

$$\begin{aligned} R^{g,g}_{\underline{tt}} R_{\underline{tt}} R^{g,g}_{\underline{tt}} =& \begin{bmatrix} \mathbb{I} & 0 \\ -R^{\dagger}_{\underline{ll}} R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{\dagger}_{\underline{m}_D \underline{m}_D} & 0 \\ 0 & R^{\dagger}_{\underline{ll}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D \underline{l}} R^{\dagger}_{\underline{ll}} \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & R_{\underline{e}_D \underline{l}} R^{\dagger}_{\underline{ll}} \\ 0 & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \Sigma_{\underline{m}_D \underline{m}_D} & 0 \\ 0 & R_{\underline{ll}} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ R^{\dagger}_{\underline{ll}} R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ -R^{\dagger}_{\underline{ll}} R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{\dagger}_{\underline{m}_D \underline{m}_D} & 0 \\ 0 & R^{\dagger}_{\underline{ll}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D \underline{l}} R^{\dagger}_{\underline{ll}} \\ 0 & \mathbb{I} \end{bmatrix}, \end{aligned} \tag{41}$$

can be simplified using

$$\begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D \underline{l}} R^{\dagger}_{\underline{ll}} \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & R_{\underline{e}_D \underline{l}} R^{\dagger}_{\underline{ll}} \\ 0 & \mathbb{I} \end{bmatrix} = \mathbb{I} \tag{42a}$$

$$\begin{bmatrix} \mathbb{I} & 0 \\ R^{\dagger}_{\underline{ll}} R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ -R^{\dagger}_{\underline{ll}} R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} = \mathbb{I}. \tag{42b}$$

Hence, $R^{g,g}_{\underline{tt}} R_{\underline{tt}} R^{g,g}_{\underline{tt}}$ can be rewritten as

$$\begin{aligned} R^{g,g}_{\underline{tt}} R_{\underline{tt}} R^{g,g}_{\underline{tt}} =& \begin{bmatrix} \mathbb{I} & 0 \\ -R^{\dagger}_{\underline{ll}} R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \Sigma^{\dagger}_{\underline{m}_D \underline{m}_D} \Sigma_{\underline{m}_D \underline{m}_D} \Sigma^{\dagger}_{\underline{m}_D \underline{m}_D} & 0 \\ 0 & R^{\dagger}_{\underline{ll}} R_{\underline{ll}} R^{\dagger}_{\underline{ll}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D \underline{l}} R^{\dagger}_{\underline{ll}} \\ 0 & \mathbb{I} \end{bmatrix}. \end{aligned} \tag{43}$$

Given that $\Sigma^{\dagger}_{\underline{m}_D \underline{m}_D} \Sigma_{\underline{m}_D \underline{m}_D} \Sigma^{\dagger}_{\underline{m}_D \underline{m}_D} = \Sigma^{\dagger}_{\underline{m}_D \underline{m}_D}$ and that $R^{\dagger}_{\underline{ll}} R_{\underline{ll}} R^{\dagger}_{\underline{ll}} = R^{\dagger}_{\underline{ll}}$, as per the definition of the pseudo-inverse, $R^{g,g}_{\underline{tt}} R_{\underline{tt}} R^{g,g}_{\underline{tt}} = R^{g,g}_{\underline{tt}}$.

Correspondingly, $R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}}$ can be decomposed as

$$R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} = \begin{bmatrix} \mathbb{I} & R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{\{\mu\}}_{\underline{mm}} & 0 \\ 0 & \mu R_{\underline{tt}} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix}, \tag{44}$$

This decomposition relies on the property $R^{g,g}_{\underline{tt}} R_{\underline{tt}} R^{g,g}_{\underline{tt}} = R^{g,g}_{\underline{tt}}$. Additionally, it relies on the property $R_{\underline{mt}} R^{g,g}_{\underline{tt}} R_{\underline{tt}} = R_{\underline{mt}}$, as $R^{g,g}_{\underline{tt}} R_{\underline{tt}}$ is a projection matrix on the row space of $R_{\underline{tt}}$ [60, Chapter 2].

Finally, $R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}}$ can be decomposed as

$$R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} = \begin{bmatrix} \mathbb{I} & 0 \\ -R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{\{\mu\}^{\dagger}}_{\underline{mm}} & 0 \\ 0 & \frac{1}{\mu} R^{g,g}_{\underline{tt}} \end{bmatrix} \begin{bmatrix} \mathbb{I} & -R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix}. \tag{45}$$

Condition 1. $R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} = R^{\{\mu\}}_{\tilde{m}\tilde{m}}$
This condition, which is a necessary and sufficient condition for generalised inverses [41], and a necessary condition for pseudo-inverses [42], is satisfied for (15).

Proof. Using (44) and (45), $R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}}$ can be expanded as

$$\begin{aligned} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} =& \begin{bmatrix} \mathbb{I} & R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{\{\mu\}}_{\underline{mm}} & 0 \\ 0 & \mu R_{\underline{tt}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ -R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \Sigma^{\{\mu\}^{\dagger}}_{\underline{mm}} & 0 \\ 0 & \frac{1}{\mu} R^{g,g}_{\underline{tt}} \end{bmatrix} \begin{bmatrix} \mathbb{I} & -R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{\{\mu\}}_{\underline{mm}} & 0 \\ 0 & \mu R_{\underline{tt}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix}, \end{aligned} \tag{46}$$

and can be simplified using

$$\begin{bmatrix} \mathbb{I} & 0 \\ R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ -R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix} = \mathbb{I} \tag{47a}$$

$$\begin{bmatrix} \mathbb{I} & -R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} = \mathbb{I}, \tag{47b}$$

to

$$\begin{aligned} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} =& \begin{bmatrix} \mathbb{I} & R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \Sigma^{\{\mu\}}_{\underline{mm}} \Sigma^{\{\mu\}^{\dagger}}_{\underline{mm}} \Sigma^{\{\mu\}}_{\underline{mm}} & 0 \\ 0 & \mu R_{\underline{tt}} R^{g,g}_{\underline{tt}} R_{\underline{tt}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix}. \end{aligned} \tag{48}$$

As per the definition of the pseudo-inverse, $\Sigma^{\{\mu\}}_{\underline{mm}} \Sigma^{\{\mu\}^{\dagger}}_{\underline{mm}} \Sigma^{\{\mu\}}_{\underline{mm}} = \Sigma^{\{\mu\}}_{\underline{mm}}$, and $R_{\underline{tt}} R^{g,g}_{\underline{tt}} R_{\underline{tt}} = R_{\underline{tt}}$. Consequently,

$$R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} = R^{\{\mu\}}_{\tilde{m}\tilde{m}}. \tag{49}$$

■

Condition 2. $R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} = R^{\{\mu\}^{g,g}}_{\tilde{m}\tilde{m}}$
This condition, which is a necessary condition for pseudo-inverses [42], is satisfied for (15).

Proof. Using (44) and (45), $R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}}$ can be expanded as

$$\begin{aligned} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}}_{\underline{\tilde{m}\tilde{m}}} R^{\{\mu\}^{g,g}}_{\underline{\tilde{m}\tilde{m}}} =& \begin{bmatrix} \mathbb{I} & 0 \\ -R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{\{\mu\}^{\dagger}}_{\underline{mm}} & 0 \\ 0 & \frac{1}{\mu} R^{g,g}_{\underline{tt}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & -R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{I} & R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \Sigma^{\{\mu\}}_{\underline{mm}} & 0 \\ 0 & \mu R_{\underline{tt}} \end{bmatrix} \begin{bmatrix} \mathbb{I} & 0 \\ R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ -R^{g,g}_{\underline{tt}} R_{\underline{tm}} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \Sigma^{\{\mu\}^{\dagger}}_{\underline{mm}} & 0 \\ 0 & \frac{1}{\mu} R^{g,g}_{\underline{tt}} \end{bmatrix} \cdot \\ & \begin{bmatrix} \mathbb{I} & -R_{\underline{mt}} R^{g,g}_{\underline{tt}} \\ 0 & \mathbb{I} \end{bmatrix}, \end{aligned} \tag{50}$$

and can be simplified using

$$\begin{bmatrix} \mathbb{I} & -R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix} = \mathbb{I} \tag{51a}$$

$$\begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix} = \mathbb{I}, \tag{51b}$$

to

$$\begin{aligned} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} =& \begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix}\cdot \\ & \begin{bmatrix} \Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}}\Sigma_{\underline{mm}}^{\{\mu\}}\Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}} & 0 \\ 0 & \frac{1}{\mu}R_{\underline{tt}}^{g,g}R_{\underline{tt}}R_{\underline{tt}}^{g,g} \end{bmatrix}\cdot \\ & \begin{bmatrix} \mathbb{I} & -R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix}. \end{aligned} \tag{52}$$

As per the definition of the pseudo-inverse, $\Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}}\Sigma_{\underline{mm}}^{\{\mu\}}\Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}} = \Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}}$, and $R_{\underline{tt}}^{g,g}R_{\underline{tt}}R_{\underline{tt}}^{g,g} = R_{\underline{tt}}^{g,g}$. Consequently,

$$R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} = R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}. \tag{53}$$

■

Condition 3. $\left(R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}\right)^H = R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$
This condition, which is a necessary condition for pseudo-inverses [42], is only satisfied for (15) when $\text{rank}(\Sigma_{\underline{mm}}^{\{\mu\}}) = M$ and $\text{rank}(\Sigma_{\underline{m}_D\underline{m}_D}) = ML_m$.

Proof. Using (44) and (45), $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$ can be expanded as

$$\begin{aligned} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} =& \begin{bmatrix} \mathbb{I} & R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix}\begin{bmatrix} \Sigma_{\underline{mm}}^{\{\mu\}} & 0 \\ 0 & \mu R_{\underline{tt}} \end{bmatrix}\cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix}\cdot \\ & \begin{bmatrix} \Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}} & 0 \\ 0 & \frac{1}{\mu}R_{\underline{tt}}^{g,g} \end{bmatrix}\begin{bmatrix} \mathbb{I} & -R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix}, \end{aligned} \tag{54}$$

and can be simplified using

$$\begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix} = \mathbb{I}, \tag{55}$$

to

$$\begin{aligned} & R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} = \\ & \begin{bmatrix} \Sigma_{\underline{mm}}^{\{\mu\}}\Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}} & -\Sigma_{\underline{mm}}^{\{\mu\}}\Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}}R_{\underline{mt}}R_{\underline{tt}}^{g,g} + R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & R_{\underline{tt}}R_{\underline{tt}}^{g,g} \end{bmatrix}. \end{aligned} \tag{56}$$

It is seen that $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$ is only Hermitian when $\Sigma_{\underline{mm}}^{\{\mu\}}\Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}} = \mathbb{I}$, i.e., $\text{rank}(\Sigma_{\underline{mm}}^{\{\mu\}}) = M$, and when $R_{\underline{tt}}R_{\underline{tt}}^{g,g}$ is Hermitian. This latter condition can be further simplified as follows. Using (36) and (37), $R_{\underline{tt}}R_{\underline{tt}}^{g,g}$ can be expanded as

$$\begin{aligned} R_{\underline{tt}}R_{\underline{tt}}^{g,g} =& \begin{bmatrix} \mathbb{I} & R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger} \\ 0 & \mathbb{I} \end{bmatrix}\begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D} & 0 \\ 0 & R_{\underline{ll}} \end{bmatrix}\cdot \\ & \begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} & \mathbb{I} \end{bmatrix}\cdot \\ & \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D}^{\dagger} & 0 \\ 0 & R_{\underline{ll}}^{\dagger} \end{bmatrix}\begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger} \\ 0 & \mathbb{I} \end{bmatrix}, \end{aligned} \tag{57}$$

and can be simplified using

$$\begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} & \mathbb{I} \end{bmatrix} = \mathbb{I}, \tag{58}$$

to

$$\begin{aligned} & R_{\underline{tt}}R_{\underline{tt}}^{g,g} \\ & = \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D}\Sigma_{\underline{m}_D\underline{m}_D}^{\dagger} & -\Sigma_{\underline{m}_D\underline{m}_D}\Sigma_{\underline{m}_D\underline{m}_D}^{\dagger}R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger} + R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger} \\ 0 & R_{\underline{ll}}R_{\underline{ll}}^{\dagger} \end{bmatrix}. \end{aligned} \tag{59}$$

It is seen that $R_{\underline{tt}}R_{\underline{tt}}^{g,g}$ is only Hermitian when $\Sigma_{\underline{m}_D\underline{m}_D}\Sigma_{\underline{m}_D\underline{m}_D}^{\dagger} = \mathbb{I}$, i.e., $\text{rank}(\Sigma_{\underline{m}_D\underline{m}_D}) = ML_m$.

In conclusion, only when $\text{rank}(\Sigma_{\underline{mm}}^{\{\mu\}}) = M$ and $\text{rank}(\Sigma_{\underline{m}_D\underline{m}_D}) = ML_m$, $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$ can be rewritten as

$$R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} = \begin{bmatrix} \mathbb{I} & 0 \\ 0 & R_{\underline{ll}}R_{\underline{ll}}^{\dagger} \end{bmatrix}, \tag{60}$$

which is indeed Hermitian. From (60) it can also be observed that $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}$ corresponds to a right inverse when additionally $\text{rank}(R_{\underline{ll}}) = LL_l$. ■

Condition 4. $\left(R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}\right)^H = R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}$
This condition, which is a necessary condition for pseudo-inverses [42], is only satisfied for (15) when $\text{rank}(\Sigma_{\underline{mm}}^{\{\mu\}}) = M$ and $\text{rank}(\Sigma_{\underline{m}_D\underline{m}_D}) = ML_m$.

Proof. Using (44) and (45), $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}$ can be expanded as

$$\begin{aligned} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} =& \begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix}\begin{bmatrix} \Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}} & 0 \\ 0 & \frac{1}{\mu}R_{\underline{tt}}^{g,g} \end{bmatrix}\cdot \\ & \begin{bmatrix} \mathbb{I} & -R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix}\cdot \\ & \begin{bmatrix} \Sigma_{\underline{mm}}^{\{\mu\}} & 0 \\ 0 & \mu R_{\underline{tt}} \end{bmatrix}\begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{tt}}^{g,g}R_{\underline{tm}} & \mathbb{I} \end{bmatrix}, \end{aligned} \tag{61}$$

and can be simplified using

$$\begin{bmatrix} \mathbb{I} & -R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & R_{\underline{mt}}R_{\underline{tt}}^{g,g} \\ 0 & \mathbb{I} \end{bmatrix} = \mathbb{I} \tag{62}$$

to

$$\begin{aligned} & R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}} R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}} = \\ & \begin{bmatrix} \Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}}\Sigma_{\underline{mm}}^{\{\mu\}} & 0 \\ -R_{\underline{tt}}^{g,g}R_{\underline{tm}}\Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}}\Sigma_{\underline{mm}}^{\{\mu\}} + R_{\underline{tt}}^{g,g}R_{\underline{tm}} & R_{\underline{tt}}^{g,g}R_{\underline{tt}} \end{bmatrix}. \end{aligned} \tag{63}$$

It is seen that $R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}^{g,g}}R_{\underline{\tilde{m}\tilde{m}}}^{\{\mu\}}$ is only Hermitian when $\Sigma_{\underline{mm}}^{\{\mu\}^{\dagger}}\Sigma_{\underline{mm}}^{\{\mu\}} = \mathbb{I}$, i.e., $\text{rank}(\Sigma_{\underline{mm}}^{\{\mu\}}) = M$, and when $R_{\underline{tt}}^{g,g}R_{\underline{tt}}$ is Hermitian. This latter condition can be further simplified as follows. Using (36) and (37), $R_{\underline{tt}}^{g,g}R_{\underline{tt}}$ can be expanded as

$$\begin{aligned} R_{\underline{tt}}^{g,g}R_{\underline{tt}} =& \begin{bmatrix} \mathbb{I} & 0 \\ -R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} & \mathbb{I} \end{bmatrix}\begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D}^{\dagger} & 0 \\ 0 & R_{\underline{ll}}^{\dagger} \end{bmatrix}\cdot \\ & \begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger} \\ 0 & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & R_{\underline{e}_D\underline{l}}R_{\underline{ll}}^{\dagger} \\ 0 & \mathbb{I} \end{bmatrix}\cdot \\ & \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D} & 0 \\ 0 & R_{\underline{ll}} \end{bmatrix}\begin{bmatrix} \mathbb{I} & 0 \\ R_{\underline{ll}}^{\dagger}R_{\underline{le}_D} & \mathbb{I} \end{bmatrix}, \end{aligned} \tag{64}$$

and can be simplified using

$$\begin{bmatrix} \mathbb{I} & -R_{\underline{e}_D\underline{l}}R_{\underline{l}\underline{l}}^{\dagger} \\ 0 & \mathbb{I} \end{bmatrix}\begin{bmatrix} \mathbb{I} & R_{\underline{e}_D\underline{l}}R_{\underline{l}\underline{l}}^{\dagger} \\ 0 & \mathbb{I} \end{bmatrix} = \mathbb{I} \tag{65}$$

to

$$R_{\underline{t}\underline{t}}^{g,g}R_{\underline{t}\underline{t}} = \begin{bmatrix} \Sigma_{\underline{m}_D\underline{m}_D}^{\dagger}\Sigma_{\underline{m}_D\underline{m}_D} & 0 \\ -R_{\underline{l}\underline{l}}^{\dagger}R_{\underline{l}\underline{e}_D}\Sigma_{\underline{m}_D\underline{m}_D}^{\dagger}\Sigma_{\underline{m}_D\underline{m}_D} + R_{\underline{l}\underline{l}}^{\dagger}R_{\underline{l}\underline{e}_D} & R_{\underline{l}\underline{l}}^{\dagger}R_{\underline{l}\underline{l}} \end{bmatrix}. \tag{66}$$

It is seen that $R_{\underline{t}\underline{t}}^{g,g}R_{\underline{t}\underline{t}}$ is only Hermitian when $\Sigma_{\underline{m}_D\underline{m}_D}^{\dagger}\Sigma_{\underline{m}_D\underline{m}_D} = \mathbb{I}$, i.e., $\text{rank}(\Sigma_{\underline{m}_D\underline{m}_D}) = ML_m$.

In conclusion, only when $\text{rank}(\Sigma_{\underline{m}\underline{m}}^{\{\mu\}}) = M$ and $\text{rank}(\Sigma_{\underline{m}_D\underline{m}_D}) = ML_m$, $R_{\tilde{m}\tilde{m}}^{\{\mu\}^{g,g}}R_{\tilde{m}\tilde{m}}^{\{\mu\}}$ can be rewritten as

$$R_{\underline{\tilde{m}}\underline{\tilde{m}}}^{\{\mu\}^{g,g}}R_{\underline{\tilde{m}}\underline{\tilde{m}}}^{\{\mu\}} = \begin{bmatrix} \mathbb{I} & 0 \\ 0 & R_{\underline{l}\underline{l}}^{\dagger}R_{\underline{l}\underline{l}} \end{bmatrix}, \tag{67}$$

which is indeed Hermitian. From (67) it can also be observed that $R_{\underline{\tilde{m}}\underline{\tilde{m}}}^{\{\mu\}^{g,g}}$ corresponds to a left inverse when additionally $\text{rank}(R_{\underline{l}\underline{l}}) = LL_l$. ■

IX. Proof of (32)

In this section, it will be shown that $\hat{W}^{\{\mu\}}$ of (19e) corresponds to the MPDR beamformer of (32) when there is only one desired speech source ($N_s = 1$), $\Sigma_{\underline{m}\underline{m}}^{\{\mu\}^{\dagger}}$ corresponds to an inverse $\Sigma_{\underline{m}\underline{m}}^{\{\mu\}^{-1}}$, $D = 1$, and μ approaches zero. The proof closely follows [11] and [38].

Proof. As $N_s = 1$ and $D = 1$, $R_{\underline{s}\underline{s}}$ simplifies to a rank-1 matrix

$$R_{\underline{s}\underline{s}} = \sigma_{\underline{d}}\mathbf{h}\mathbf{h}^H, \tag{68}$$

with $\mathbf{h} \in \mathbb{C}^M$ collecting the acoustic transfer functions (ATFs) between the desired speech source and the microphones, and with $\sigma_{\underline{d}}$ the prefiltered PSD of the desired speech source. Consequently, $\Sigma_{\underline{m}\underline{m}}^{\{\mu\}}$ of (17b) can be decomposed as the sum of this rank-1 matrix, and the auto-correlation matrix of the interfering echo, late reverberation component of the desired speech signal, and noise, following the AEC and DR filters. Indeed,

$$R_{\underline{i}\underline{i}} = R_{\underline{r}\underline{r}} + R_{\underline{e}\underline{e}} + R_{\underline{n}\underline{n}} - R_{\underline{m}\underline{t}}R_{\underline{t}\underline{t}}^{g,g}R_{\underline{t}\underline{m}}, \tag{69}$$

with

$$\Sigma_{\underline{m}\underline{m}}^{\{\mu\}} = R_{\underline{s}\underline{s}} + \mu R_{\underline{i}\underline{i}}. \tag{70}$$

When $\Sigma_{\underline{m}\underline{m}}^{\{\mu\}^{\dagger}}$ corresponds to an inverse $\Sigma_{\underline{m}\underline{m}}^{\{\mu\}^{-1}}$, this inverse can be specified using the Sherman-Morrison formula [58]. The filter $\hat{W}^{\{\mu\}}$ can subsequently be rewritten as a beamformer followed by a single-channel post-filter. Indeed, applying the Sherman-Morrison formula [58] yields

$$\begin{aligned} \hat{W}^{\{\mu\}} &= \left(\sigma_{\underline{d}}^2\mathbf{h}\mathbf{h}^H + \mu R_{\underline{i}\underline{i}}\right)^{-1}\sigma_{\underline{d}}^2\mathbf{h}\mathbf{h}^H\mathbf{t}_q && \text{(71a)} \\ &= \left(\frac{1}{\mu}R_{\underline{i}\underline{i}}^{-1} - \frac{\sigma_{\underline{d}}^2}{\mu^2}\frac{R_{\underline{i}\underline{i}}^{-1}\mathbf{h}\mathbf{h}^H R_{\underline{i}\underline{i}}^{-1}}{1+\frac{\sigma_{\underline{d}}^2}{\mu}\mathbf{h}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}}\right)\sigma_{\underline{d}}^2\mathbf{h}\mathbf{h}^H\mathbf{t}_q && \text{(71b)} \\ &= \frac{\frac{\sigma_{\underline{d}}^2}{\mu}R_{\underline{i}\underline{i}}^{-1}\mathbf{h}\mathbf{h}^H\mathbf{t}_q}{1+\frac{\sigma_{\underline{d}}^2}{\mu}\mathbf{h}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}}. && \text{(71c)} \end{aligned}$$

In (71c), $\mathbf{h}$ can also be replaced with the following expression

$$\mathbf{h} = \underbrace{\frac{\mathbf{h}}{\mathbf{t}_q^{\top}\mathbf{h}}}_{\triangleq\mathbf{h}_{\text{rel}}}\cdot\mathbf{t}_q^{\top}\mathbf{h}, \tag{72}$$

where $\mathbf{h}_{\text{rel}} \in \mathbb{C}^M$ collects the relative transfer functions (RTFs) from the reference microphone q to the other microphones. With (72), (71c) can be rewritten as

$$\hat{W}^{\{\mu\}} = \underbrace{\frac{\sigma_{\underline{d}}^2|\mathbf{t}_q^{\top}\mathbf{h}|^2}{\sigma_{\underline{d}}^2|\mathbf{t}_q^{\top}\mathbf{h}|^2 + \mu\left(\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}\right)^{-1}}}_{\text{Single-channel post-filter}}\underbrace{\frac{R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}{\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}}_{\text{MVDR}}, \tag{73}$$

which corresponds to a minimum variance distortionless response (MVDR) beamformer, followed by a single-channel post-filter. As in [11], when μ approaches zero, the single-channel post-filter disappears, such that only the MVDR beamformer remains as

$$\hat{W}^{\{\mu\to 0\}} = \frac{R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}{\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}. \tag{74}$$

As in [38], this MVDR beamformer, requiring the inverse of $R_{\underline{i}\underline{i}}$, can be interchanged with an MPDR beamformer, requiring the inverse of $\Sigma_{\underline{m}\underline{m}}^{\{\mu=1\}}$. Indeed, as in [38], and using the Sherman-Morrison formula [58], $\Sigma_{\underline{m}\underline{m}}^{\{\mu=1\}^{-1}}\mathbf{h}_{\text{rel}}$ can be rewritten as

$$\begin{aligned} \Sigma_{\underline{m}\underline{m}}^{\{\mu=1\}^{-1}}\mathbf{h}_{\text{rel}} &= (R_{\underline{i}\underline{i}} + \sigma_{\underline{d}}^2|\mathbf{t}_q^{\top}\mathbf{h}|^2\mathbf{h}_{\text{rel}}\mathbf{h}_{\text{rel}}^H)^{-1}\mathbf{h}_{\text{rel}} && \text{(75a)} \\ &= \left(R_{\underline{i}\underline{i}}^{-1} - \frac{|\mathbf{t}_q^{\top}\mathbf{h}|^2\sigma_{\underline{d}}^2 R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}}{1+\sigma_{\underline{d}}^2|\mathbf{t}_q^{\top}\mathbf{h}|^2\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}\right)\mathbf{h}_{\text{rel}} && \text{(75b)} \\ &= \frac{R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}{1+\sigma_{\underline{d}}^2|\mathbf{t}_q^{\top}\mathbf{h}|^2\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}, && \text{(75c)} \end{aligned}$$

and $\mathbf{h}_{\text{rel}}^H\Sigma_{\underline{m}\underline{m}}^{\{\mu=1\}^{-1}}\mathbf{h}_{\text{rel}}$ can similarly be rewritten as

$$\mathbf{h}_{\text{rel}}^H\Sigma_{\underline{m}\underline{m}}^{\{\mu=1\}^{-1}}\mathbf{h}_{\text{rel}} = \frac{\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}{1+\sigma_{\underline{d}}^2|\mathbf{t}_q^{\top}\mathbf{h}|^2\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}. \tag{76}$$

As a result, the MVDR beamformer of (73) can be equivalently implemented using an MPDR beamformer, i.e.,

$$\begin{aligned} \hat{W}^{\{\mu\to 0\}} &= \frac{R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}}{\mathbf{h}_{\text{rel}}^H R_{\underline{i}\underline{i}}^{-1}\mathbf{h}_{\text{rel}}} && \text{(77a)} \\ &= \frac{\Sigma_{\underline{m}\underline{m}}^{\{\mu=1\}^{-1}}\mathbf{h}_{\text{rel}}}{\mathbf{h}_{\text{rel}}^H\Sigma_{\underline{m}\underline{m}}^{\{\mu=1\}^{-1}}\mathbf{h}_{\text{rel}}}. && \text{(77b)} \end{aligned}$$

It is notated that it is also possible to use $\Sigma_{\underline{m}\underline{m}}^{\{\mu'\}}$, with a weighting $\mu' \neq \mu$, rather than to use $\Sigma_{\underline{m}\underline{m}}^{\{\mu=1\}}$. This has been omitted for clarity as an MPDR beamformer with such weighting is typically not considered [38]. ■

REFERENCES

[11] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio Speech Language Process.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.

[38] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*. John Wiley & Sons, Ltd, Jan. 2010, ch. 9, pp. 269–302.

[41] M. James, "The generalised inverse," *Math. Gaz.*, vol. 62, no. 420, pp. 109–114, Jun. 1978.

[42] R. Penrose, "A generalized inverse for matrices," *Math. Proc. Cambridge Philos. Soc.*, vol. 51, no. 3, pp. 406–413, Jul. 1955.

[58] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Stat.*, vol. 21, no. 1, pp. 124–127, Mar. 1950.

[59] G. H. Golub and C. F. Van Loan, *Matrix computations*, 4th ed. Baltimore, MD, USA: JHU press, Feb. 2013.

[60] A. Ben-Israel and T. N. Greville, *Generalized inverses: Theory and applications*, 2nd ed. New York, NY, USA: Springer New York, Jun. 2003.