A comparison between overlap-save and weighted overlap-add filter banks for multi-channel Wiener filter based noise reduction

Santiago Ruiz, Thomas Dietzen, Toon van Waterschoot and Marc Moonen Dept. of Electrical Engineering, ESAT-STADIUS

KU Leuven

Leuven, Belgium

Email: {santiago.ruiz,thomas.dietzen,toon.vanwaterschoot,marc.moonen}@esat.kuleuven.be

Abstract—A comparison is provided between multi-channel Wiener filter (MWF) implementations for noise reduction (NR) using overlap-save (OLS) and weighted overlap-add (WOLA) filter banks. Simulations are used to show the effect of constraining the filters in an OLS-based implementation as well as differences in the estimated correlation matrices and NR filters using different filter banks. Overall, the WOLA-based implementation provides better NR performance in comparison to the OLS-based implementation. The rectangular analysis window and constraining the filter in the OLS-based implementation deteriorates the performance of the MWF.

Index Terms—Noise reduction, multi-channel Wiener filter, overlap-save, weighted overlap-add, filter banks.

I. INTRODUCTION

Noise reduction (NR) is used to enhance a desired speech signal in today's speech communication systems, including, but not limited to, hands-free telephony, hearing aids, automatic speech recognition and teleconferencing. By using multiple microphones it is possible to exploit spatial characteristics of an acoustic scenario. Such multi-channel NR generally results in a better performance compared to single-channel NR, particularly when the desired speech and noise sources are spatially separated.

A widely used NR technique is the multi-channel Wiener filter (MWF) [1]. Adaptive beamforming techniques, such as the minimum variance distortionless response (MVDR) beamformer are also commonly used [2]. Such techniques usually rely on an overlap-save (OLS) or weighted overlapadd (WOLA) filter bank to perform efficient time-domain or frequency-domain (subband) filtering, respectively [3]–[5]. WOLA filter banks require analysis and synthesis windows which reduce the effects of circular convolution and improve side-lobe rejection in the frequency domain. OLS filter banks are implemented with a rectangular analysis window [4]. In

OLS filter banks, constrained filters have been implemented, where the filtering operation performed in the frequencydomain is constrained to correspond exactly to a time-domain filtering operation [4], [6]. However, the additional constraints can introduce dependencies across frequency bins [7]. Such effects are usually not a problem in some applications, such as acoustic echo cancellation [8]-[10]. Alternatively, unconstrained filters have been used in OLS filter banks, which cannot fully prevent circular convolution effects [4], [11], [12]. Multiple WOLA-based MWF implementations have been reported in the literature, however a comparison between WOLA- and OLS-based MWF has not been reported. In this paper, the performance of MWF based NR using WOLA, unconstrained OLS (uOLS) and constrained OLS (cOLS) filter banks is studied. Simulations are used to show the effect of constraining the filter in an OLS-based implementation as well as differences in the estimated correlation matrices and NR filters using different filter banks. Overall, the WOLAbased implementation provides better NR performance in comparison to the OLS-based implementation. The rectangular analysis window and constraining the filter in the OLS-based implementation deteriorates the performance of the MWF. The paper is organized as follows. In Section II the signal model is stated and the MWF is reviewed. The WOLA-, uOLS- and cOLS-based MWF implementations are described in Section III. Simulations are presented in Section IV and Section V concludes the paper.

II. SIGNAL MODEL AND MULTI-CHANNEL WIENER FILTER

An $m \times 1$ multi-microphone signal vector in the short-time Fourier transform (STFT) domain, i.e. after an OLS or WOLA analysis filter bank, as defined in Section III, is expressed as follows

$$\mathbf{y}(\kappa, l) = \mathbf{s}(\kappa, l) + \mathbf{n}(\kappa, l) \tag{1}$$

where s and n represent the desired speech and noise components, respectively, l is the time frame index, κ the frequency bin (subband) index and m the number of microphones. The

This work was carried out at the ESAT Laboratory of KU Leuven in the frame of KU Leuven Internal Fund Projects VES/19/004 and C2-16-00449 "Distributed Digital Signal Processing for Ad-hoc Wireless Local Area Audio Networking", FWO/FNRS EOS Project nr. 30452698 "MUSE-WINET - Multi-Service Wireless Network" and European Research Council under the European Union's Horizon 2020 Research and Innovation Program / ERC Consolidator Grant: SONORA (no. 773268) . The scientific responsibility is assumed by its authors.

microphone, (unknown) speech and noise correlation matrices are defined, respectively, as

$$\bar{\mathbf{R}}_{\mathbf{y}\mathbf{y}}(\kappa, l) = E\{\mathbf{y}(\kappa, l)\mathbf{y}^{H}(\kappa, l)\}$$
(2)

$$\bar{\mathbf{R}}_{\mathbf{ss}}(\kappa, l) = E\{\mathbf{s}(\kappa, l)\mathbf{s}^{H}(\kappa, l)\}$$
(3)

$$\bar{\mathbf{R}}_{\mathbf{nn}}(\kappa, l) = E\{\mathbf{n}(\kappa, l)\mathbf{n}^{H}(\kappa, l)\}$$
(4)

where $E\{\cdot\}$ is the expected value operator and $(\cdot)^H$ is the complex conjugate transpose operator. It is assumed that $\mathbf{s}(\kappa, l)$ and $\mathbf{n}(\kappa, l)$ are uncorrelated, and that $\mathbf{\bar{R}_{ss}}(\kappa, l)$ can be approximated by a rank-1 matrix if there is only one desired speech source [13].

Using the desired speech component in the first microphone as desired signal, i.e., $d(\kappa, l) = \mathbf{e}^T \mathbf{s}(\kappa, l)$ with $\mathbf{e} = [1 \ \mathbf{0}]^T$ a vector with matching dimensions that selects the first column of a matrix, the MWF is defined as the minimization of the mean squared error (MSE) between the desired signal and the filtered microphone signals, i.e., [1], [13].

$$\bar{\mathbf{w}}(\kappa, l) = \underset{\mathbf{w}}{\operatorname{arg\,min}} E\left\{ \left| d(\kappa, l) - \mathbf{w}(\kappa, l)^{H} \mathbf{y}(\kappa, l) \right|^{2} \right\}, \quad (5)$$

where the solution to this minimization problem is given by

$$\bar{\mathbf{w}}(\kappa, l) = \bar{\mathbf{R}}_{\mathbf{y}\mathbf{y}}^{-1}(\kappa, l)\bar{\mathbf{R}}_{\mathbf{ss}}(\kappa, l)\mathbf{e}.$$
(6)

The desired signal estimate is obtained by filtering the frequency-domain multi-microphone signal, i.e.,

$$\hat{d}(\kappa, l) = \bar{\mathbf{w}}^{H}(\kappa, l) \mathbf{y}(\kappa, l).$$
(7)

Given that $\bar{\mathbf{R}}_{ss}$ is not directly observable, "speech-plus-noise" and "noise-only" correlation matrices can be estimated using a voice activity detector (VAD), as follows

$$\hat{\mathbf{R}}_{\mathbf{yy}}(\kappa, l) = \lambda \hat{\mathbf{R}}_{\mathbf{yy}}(\kappa, l-1) + (1-\lambda)\mathbf{y}(\kappa, l)\mathbf{y}^{H}(\kappa, l) \quad (8)$$
$$\hat{\mathbf{R}}_{\mathrm{exc}}(\kappa, l) = \lambda \hat{\mathbf{R}}_{\mathrm{exc}}(\kappa, l-1) + (1-\lambda)\mathbf{y}(\kappa, l)\mathbf{y}^{H}(\kappa, l) \quad (9)$$

$$\mathbf{n}_{\mathbf{nn}}(k,l) = \lambda \mathbf{n}_{\mathbf{nn}}(k,l-1) + (1-\lambda)\mathbf{y}(k,l)\mathbf{y}^{-1}(k,l)$$

Here λ is a forgetting factor that is chosen according to the time variation of the signal statistics, e.g., for long-term estimates $\lambda \approx 1$ to mainly capture spatial coherence between the microphone signals. With (8) and (9), an estimate of the speech correlation matrix can be obtained based on a joint diagonalization of the matrix pencil { $\hat{\mathbf{R}}_{yy}(\kappa, l)$, $\hat{\mathbf{R}}_{nn}(\kappa, l)$ } [13], [14],

$$\hat{\mathbf{R}}_{\mathbf{yy}}(\kappa, l) = \hat{\mathbf{Q}}(\kappa, l) \hat{\boldsymbol{\Sigma}}_{\mathbf{yy}}(\kappa, l) \hat{\mathbf{Q}}^{H}(\kappa, l)$$
(10)
$$\hat{\mathbf{R}}_{\mathbf{nn}}(\kappa, l) = \hat{\mathbf{Q}}(\kappa, l) \hat{\boldsymbol{\Sigma}}_{\mathbf{nn}}(\kappa, l) \hat{\mathbf{Q}}^{H}(\kappa, l)$$

where $\hat{\mathbf{Q}}$ is an invertible matrix and $\hat{\mathbf{\Sigma}}_{yy} = \text{diag}\{\hat{\sigma}_{y_1},...,\hat{\sigma}_{y_m}\}\$ and $\hat{\mathbf{\Sigma}}_{nn} = \text{diag}\{\hat{\sigma}_{n_1},...,\hat{\sigma}_{n_m}\}\$ are diagonal matrices. The operator $\text{diag}\{\cdot\}\$ arranges the elements of its argument in a diagonal matrix. A rank-1 speech correlation matrix estimate $\hat{\mathbf{R}}_{ss}$ is then computed as

$$\hat{\mathbf{R}}_{ss}(\kappa,l) = \hat{\mathbf{Q}}(\kappa,l) \text{diag}\{\hat{\sigma}_{y_1} - \hat{\sigma}_{n_1}, 0, \dots, 0\} \hat{\mathbf{Q}}^H(\kappa,l)$$
(11)

where $\hat{\sigma}_{y_1}$ and $\hat{\sigma}_{n_1}$ are the first diagonal elements of $\hat{\Sigma}_{yy}$ and $\hat{\Sigma}_{nn}$, respectively, which correspond to the largest ratio $\hat{\sigma}_{y_i}/\hat{\sigma}_{n_i}$. Using (11), the expression in (6) then becomes

$$\hat{\mathbf{w}}(\kappa, l) = \hat{\mathbf{Q}}^{-H} \operatorname{diag} \left\{ 1 - \frac{\hat{\sigma}_{n_1}}{\hat{\sigma}_{y_1}}, 0, \dots, 0 \right\} \hat{\mathbf{Q}}^{H} \mathbf{e}, \qquad (12)$$

which can then replace $\bar{\mathbf{w}}(\kappa, l)$ in (7).

III. OLS- AND WOLA-BASED IMPLEMENTATIONS

For each implementation it is assumed that an R samples long analysis window \breve{g}_a with a 50% overlap is used to transform time-domain signals to the STFT domain, e.g., the $l^{\rm th}$ time frame of the $m^{\rm th}$ -microphone signal can be defined in the discrete-time domain as

$$\breve{y}_m(n,l) = y_m\left(n+l\frac{R}{2}\right)g_a(n) \tag{13}$$

where the time index $n \in \{0, 1, ..., R - 1\}$ and $l \in \{0, 1, ..., L - 1\}$, where *L* is the total number of time frames. The discrete Fourier transform (DFT) matrix \mathcal{F}_R of size $R \times R$ is then used to obtain the STFT representation. Similarly, synthesis windows \mathbf{g}_s are used to obtain the estimated signal. In weighted overlap-add (WOLA) filter banks, \mathbf{g}_a and \mathbf{g}_s are carefully selected window functions, i.e., Hann, squared-root Hann, Hamming, etc. Squared-root Hann windows are commonly used because they allow perfect reconstruction. The windows smooth out the effects of the circular convolution and improve the frequency selectivity. The desired signal estimate in the time domain using a WOLA filter bank is obtained by first using the inverse of the DFT matrix \mathcal{F}_R and then applying a synthesis window, as follows

$$\breve{\mathbf{d}}^{\text{WOLA}}(l) = \breve{\mathbf{G}}_{s}^{\text{WOLA}} \boldsymbol{\mathcal{F}}_{R}^{-1} \mathbf{\hat{d}}^{T}(l)$$
(14)

with

$$\hat{\mathbf{d}}(l) = \begin{bmatrix} \hat{d}(0,l) & \dots & \hat{d}(R-1,l) \end{bmatrix}$$
(15)

an $R \times 1$ vector containing all the frequency bins of the desired signal at time frame l, $\mathbf{\breve{G}}_s^{\text{WOLA}} = \text{diag}\{\mathbf{\breve{g}}_s\}$ with $\mathbf{\breve{g}}_s$ an $R \times 1$ vector containing a synthesis window. The estimated signal is obtained in the discrete-time domain by adding the L overlapping windowed frames as

$$\breve{d}(n) = \sum_{l=0}^{L-1} \hat{d}\left(n - l\frac{R}{2}, l\right) g_s(n).$$
(16)

Note that the operations in WOLA filter banks do not exactly correspond to a time-domain convolution, however the circular convolution effects are limited due to smooth windows. Alternatively, an overlap-save (OLS) filter bank can be implemented, where a rectangular window is used (i.e. \breve{g}_a is a vector of ones) and the first $\frac{R}{2}$ samples of the time-domain desired signal estimate are discarded as [4]

$$\breve{\mathbf{d}}^{\text{OLS}}(l) = \breve{\mathbf{G}}_{s}^{\text{OLS}} \boldsymbol{\mathcal{F}}_{R}^{-1} \mathbf{\hat{d}}^{T}(l)$$
(17)

where $\breve{\mathbf{G}}_{s}^{\text{OLS}} = \begin{bmatrix} \mathbf{0}_{\frac{R}{2}} & \mathbf{0}_{\frac{R}{2}} \\ \mathbf{0}_{\frac{R}{2}} & \mathbf{I}_{\frac{R}{2}} \end{bmatrix}$, $\mathbf{0}_{\frac{R}{2}}$ and $\mathbf{I}_{\frac{R}{2}}$ are all-zero and

identity matrices of size $\frac{R}{2} \times \frac{R}{2}$, respectively. This method is referred to as unconstrained-OLS (uOLS). Note that the circular convolution effects are not completely avoided because the estimated filters are not constrained, i.e., no changes to (12) are included. To avoid such effects, each NR filter can be replaced by a subspace projection whose last $\frac{R}{2}$ samples in the time domain are zero, as follows [6]

$$\hat{\mathbf{W}}_{\substack{\text{cOLS1}\\m\times R}}(l) = \left(\boldsymbol{\mathcal{F}}_{R} \mathbf{\breve{G}}_{s}^{\text{cOLS1}} \boldsymbol{\mathcal{F}}_{R}^{-1} \mathbf{\widehat{W}}^{T}(l)\right)^{T}, \quad (18)$$

with

$$\hat{\mathbf{W}}(l) = \left[\hat{\mathbf{w}}(0,l)\dots\hat{\mathbf{w}}(R-1,l)\right],\tag{19}$$

$$\hat{\mathbf{W}}_{\text{cOLS1}}(l) = \begin{bmatrix} \hat{\mathbf{w}}_{\text{cOLS1}}(0, l) \dots \hat{\mathbf{w}}_{\text{cOLS1}}(R-1, l) \end{bmatrix}$$
(20)
$$m \times R$$

and $\breve{\mathbf{G}}_{s}^{\text{cOLS1}} = \begin{bmatrix} \mathbf{I}_{\frac{R}{2}} & \mathbf{0}_{\frac{R}{2}} \\ \mathbf{0}_{\frac{R}{2}} & \mathbf{0}_{\frac{R}{2}} \end{bmatrix}$. The columns in (19) and (20) are the filters per frequency bin which are used to estimate the

desired signal as in (7). The projection is performed so that the operation exactly corresponds to a linear convolution in the time domain. This method is referred to as constrained-OLS (cOLS1). Alternatively to (18), the constrained filter $\hat{\mathbf{w}}_{\text{cOLS2}}(l)$ can be computed as

$$\hat{\mathbf{W}}_{\substack{\text{cOLS2}\\m\times R}}(l) = \left(\boldsymbol{\mathcal{F}}_{R} \begin{bmatrix} \boldsymbol{\mathcal{F}}_{\frac{R}{2}}^{-1} \mathbf{C}_{\frac{R}{2} \times R} \hat{\mathbf{W}}^{T}(l) \\ \mathbf{0}_{\frac{R}{2} \times m} \end{bmatrix} \right)^{T}$$
(21)

where $\mathbf{C}_{\frac{R}{2} \times R}$ is a $\frac{R}{2} \times R$ matrix that selects every second frequency bin (0, 2, 4, ...) of $\hat{\mathbf{W}}(l)$ in (19) and $\hat{\mathbf{W}}_{\text{cOLS2}}(l)$ is defined similarly to (20). This approach keeps the selected frequency bin values and the remaining ones are interpolated by the inverse DFT and DFT matrices, based on the constraint in the time domain, i.e., the last $\frac{R}{2}$ samples of the timedomain filter must be zero. The filters obtained with the cOLSbased implementations correspond to an exact time domain convolution, however constraining the filter changes the filter coefficients, hence they are not optimal per frequency bin anymore, i.e., they are not the optimal solution to (5).

Table I shows a summary of the methods described in this section.

IV. SIMULATIONS

A. Scenario description

The performance of the MWF implementations using WOLA, uOLS and cOLS was assessed in an scenario with a 3microphone linear array placed in a room in front of a desired source and a localized noise source. Microphone signals were generated using room impulse responses (RIR) simulated with the randomized image method described in [15]. The sampling frequency was 16 kHz, the length of the RIRs was set to 512 samples and the room's reverberation time was $T_{60} \approx 0.11$ s. The localized noise source played back a white noise signal which was then convolved with the RIRs for each microphone. The following scenarios are defined based on the signals used:

- Scenario 1 The desired source played back an ON-OFF speech signal convolved with the corresponding RIR for each microphone and a localized source was included.
- Scenario 2 The desired source played back an ON-OFF white noise signal convolved with the corresponding RIR for each microphone and a localized source was included.

• Scenario 3 The desired source played back an ON-OFF speech signal convolved with the corresponding RIR for each microphone and a localized source was included. The desired speech source is located at 135° with respect to the centre of the array and a noise source at 25°.

B. Correlation matrices estimates

The estimates of the power spectral density for each channel, i.e., the diagonal elements of the correlation matrices $\hat{\mathbf{R}}_{yy}$ and $\hat{\mathbf{R}}_{nn}$ for all frequency bins, are shown in Fig. 1 for different implementations using Scenario 1. A significant difference is observed above 4 kHz where in the OLS-based implementations an almost flat response is found as opposed to the peaks observed in the WOLA-based implementation. The differences can be explained by the high amplitude of the side lobes in the rectangular analysis window used by the OLS analysis filter bank.

C. Estimated filters

The estimated filters for the first channel using different implementations are shown in Fig. 2 (similar results were obtained for the remaining channels, but are not shown for brevity) using Scenario 2. Differences in the filters are mainly observed below 250 Hz and above 4 kHz. In both cases the window function in the WOLA-based implementation provides better frequency selectivity, and this can be seen in the magnitude of the frequency response of the filters. The flat responses above 4 kHz in the OLS-based MWF implementation are likely due to the poor resolution of the rectangular analysis window, which causes estimation errors in the correlation matrices. This leads to no differences in magnitude in this frequency range, which causes the magnitude response of the estimated filters to be flat. It is observed that constraining the filters in the OLS-based implementations smooths the magnitude of their frequency responses.

D. Directivity patterns, DFT size and iSNR

The SNR improvement (Δ SNR = oSNR - iSNR) was computed for MWF implementations using WOLA, uOLS, cOLS and cOLS2 with different DFT sizes. The input and output SNRs are denoted by iSNR and oSNR, respectively. Fig. 3 shows the Δ SNR for different MWF implementations using Scenario 1. For comparison the localized white noise source was removed and uncorrelated white noise was added to the microphone signals. In both scenarios the WOLAbased implementation outperforms the uOLS- and cOLS-based implementations. The performance of the WOLA- and uOLSbased implementations improves when correlated noise is used, whereas the performance of the cOLS-based implementations does not improve. In Fig. 4A the oSNR is shown when different iSNRs are used. The filter length was set to 512 samples, which is equal to the RIRs length. It can be seen that the WOLA-based implementation outperforms all other implementations for all iSNR.

WOLA	uOLS	cOLS1	cOLS2
Obtain STFT representation using the analysis window			
$\check{\mathbf{g}}_a^{ ext{WOLA}}$	$old {f g}_a^{ m OLS}$	$old {f g}_a^{ m OLS}$	$\breve{\mathbf{g}}_a^{ ext{OLS}}$
Update $\hat{\mathbf{R}}_{\mathbf{yy}}(\kappa, l)$ and $\hat{\mathbf{R}}_{\mathbf{nn}}(\kappa, l)$ using (8) and (9)			
Compute $\hat{\mathbf{w}}(\kappa, l)$ based on GEVD of $\{\hat{\mathbf{R}}_{\mathbf{yy}}(\kappa, l), \hat{\mathbf{R}}_{\mathbf{nn}}(\kappa, l)\}$ by using (12)			
$\mathbf{\hat{w}}_{\mathrm{WOLA}}(\kappa,l) = \mathbf{\hat{w}}(\kappa,l)$	$\mathbf{\hat{w}}_{\rm uOLS}(\kappa,l) = \mathbf{\hat{w}}(\kappa,l)$	$\begin{split} \mathbf{\hat{W}}_{\text{cOLS1}}(l) &= \\ \left(\boldsymbol{\mathcal{F}}_{R} \mathbf{\breve{G}}_{s}^{\text{cOLS1}} \boldsymbol{\mathcal{F}}_{R}^{-1} \mathbf{\hat{W}}^{T}(l) \right)^{T} \end{split}$	$ \begin{split} \hat{\mathbf{W}}_{\text{cOLS2}}(l) &= \\ & \left(\boldsymbol{\mathcal{F}}_{R} \begin{bmatrix} \boldsymbol{\mathcal{F}}_{\frac{R}{2} \times R}^{-1} \mathbf{C}_{\frac{R}{2} \times R} \hat{\mathbf{W}}^{T}(l) \\ & 0_{\frac{R}{2} \times m} \end{bmatrix} \right)^{T} \end{split} $
$\hat{d}(\kappa, l) = \hat{\mathbf{w}}_{\text{WOLA}}^{H}(\kappa, l) \mathbf{y}(\kappa, l)$	$\hat{d}(\boldsymbol{\kappa},l) = \hat{\mathbf{w}}_{\mathrm{uOLS}}^{H}(\boldsymbol{\kappa},l) \mathbf{y}(\boldsymbol{\kappa},l)$	$\hat{d}(\kappa, l) = \hat{\mathbf{w}}_{\text{cOLS1}}^{H}(\kappa, l) \mathbf{y}(\kappa, l)$	$\hat{d}(\kappa, l) = \hat{\mathbf{w}}_{\text{cOLS2}}^{H}(\kappa, l) \mathbf{y}(\kappa, l)$
$\check{\mathbf{d}}^{\mathrm{WOLA}}(l) = \check{\mathbf{G}}_{s}^{\mathrm{WOLA}} \boldsymbol{\mathcal{F}}_{R}^{-1} \hat{\mathbf{d}}(l)$	$\breve{\mathbf{d}}^{\mathrm{uOLS}}(l) = \breve{\mathbf{G}}_s^{\mathrm{OLS}} \boldsymbol{\mathcal{F}}_R^{-1} \mathbf{\hat{d}}(l)$	$\breve{\mathbf{d}}^{\text{cOLS1}}(l) = \breve{\mathbf{G}}_s^{\text{OLS}} \boldsymbol{\mathcal{F}}_R^{-1} \mathbf{\hat{d}}(l)$	$\breve{\mathbf{d}}^{\text{cOLS2}}(l) = \breve{\mathbf{G}}_s^{\text{OLS}} \boldsymbol{\mathcal{F}}_R^{-1} \mathbf{\hat{d}}(l)$

TABLE I: Summary of the WOLA-, uOLS-, cOLS1- and cOLS2-based MWF implementations to obtain the NR filters and the desired speech signal.



Fig. 1: Diagonal elements of the estimated correlation matrices $\hat{\mathbf{R}}_{yy}$ and $\hat{\mathbf{R}}_{nn}$ for all frequency bins for WOLA- and OLS-based implementations using Scenario 1.



Fig. 2: Filters applied to the first channel using WOLA-, uOLS-, cOLS1- and cOLS2-based implementations using Scenario 2.



Fig. 3: Δ SNR against DFT size using Scenario 1 when **A** uncorrelated noise is used in the microphone signals. **B** localized white noise source is used in the microphone signals. The length of the simulated RIRs is $L_f = 512$.

Directivity patterns of the estimated filters were computed for each implementation as

$$\mathbf{H}(\kappa,\theta) = \mathbf{w}^{H}(\kappa)\mathbf{a}(\kappa,\theta) \tag{22}$$

where $\mathbf{a}(\kappa, \theta)$ is an $m \times 1$ vector that contains the acoustic transfer functions evaluated in frequency bin κ from a source located at an angle θ to each microphone in the array. The average of the magnitudes over frequency bins κ for WOLA-, uOLS- and cOLS1-based implementations are shown in Fig. 4B for Scenario 3. It is observed that the WOLA-based implementation is the most effective in terms of reducing the contribution from the noise source, where a low magnitude response in the noise source direction is observed. The uOLS-based implementation's directivity pattern provides the largest gain for the desired speech source but its rejection of the noise source contribution is similar to that of the cOLS1-based implementation.

V. CONCLUSIONS

The WOLA-based MWF implementation outperforms the uOLS- and cOLS-based implementations in terms of SNR and its directivity pattern rejects better the contributions from the localized noise source. The relative performance of the methods does not depend on the iSNR used. The rectangular analysis window used in the uOLS-based implementation prevents this method from achieving a similar performance to the WOLA-based implementation. The cOLS-based implementation performs poorer than the WOLA- and uOLS-based implementations in terms of NR. The MWF obtained with the WOLA-based implementation is optimal in each frequency bin which is not the case for cOLS-based implementations as the filter coefficients are changed. Similarly, the uOLS-based implementation is also optimal in each frequency bin but the use of a rectangular analysis window harms its performance.



Fig. 4: A oSNR for different iSNRs when a localized white noise source is used in the microphone signals. The filter length was 512 samples. B Directivity patterns of the WOLA-, uOLS- and cOLS1-based MWF implementations using Scenario 3. The dashed lines indicate the direction of the desired speech and noise sources.

REFERENCES

- J. Benesty, J. Chen, Y. A. Huang, and S. Doclo, "Study of the Wiener filter for noise reduction," in *Speech Enhancement*, pp. 9–41. Springer, 2005.
- [2] M. Souden, J. Benesty, and S. Affes, "A study of the LCMV and MVDR noise reduction filters," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4925–4935, 2010.
- [3] R. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 1, pp. 99–102, 1980.
- [4] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, 1992.
- [5] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, 2007.
- [6] W. Kellermann and H. Buchner, "Wideband algorithms versus narrowband algorithms for adaptive filtering in the DFT domain," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, 2003, vol. 2, pp. 1278–1282.
- [7] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Acoust.*, *Speech, Signal Process.*, vol. 15, no. 4, pp. 1305–1319, 2007.
- [8] G. Enzner, A model based optimum filtering approach to acoustic echo control: Theory and practice, Ph.D. thesis, RWTH Aachen University, 2006.
- [9] F. Kuech, E. Mabande, and G. Enzner, "State-space architecture of the partitioned-block-based acoustic echo controller," in *Proc. 2014 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '14)*, 2014.
- [10] M. L. Valero, E. Mabande, and E. A. P. Habets, "A state-space partitioned-block adaptive filter for echo cancellation using inter-band correlations in the Kalman gain computation," in *Proc. 2015 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '15)*, 2015.
- [11] D. Mansour and A. Gray, "Unconstrained frequency-domain adaptive filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 5, pp. 726–734, 1982.
- [12] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140–1156, 2006.
- [13] 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 Trans. Audio Speech Lang. Process.*, vol. 22, no. 4, pp. 785–799, 2014.
- [14] F. Jabloun and B. Champagne, "Signal subspace techniques for speech enhancement," in *Speech Enhancement*, pp. 135–159. Springer, 2005.
- [15] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Trans. Audio, Speech, Language Process*, vol. 23, no. 4, pp. 774–786, 2015.