CENTRALIZED CASCADE MULTI-CHANNEL NOISE REDUCTION AND ACOUSTIC FEEDBACK CANCELLATION IN A WIRELESS ACOUSTIC SENSOR AND ACTUATOR NETWORK

Santiago Ruiz, Toon van Waterschoot and Marc Moonen

KU Leuven
Dept. of Electrical Engineering ESAT-STADIUS
Leuven, Belgium

ABSTRACT

This paper presents a centralized cascade multi-channel noise reduction (NR) and acoustic feedback cancellation (AFC) algorithm for speech applications in a wireless acoustic sensor and actuator network (WASAN). The algorithm consists of a multi-channel Wiener filter (MWF) based NR stage, where M microphone and L loudspeaker signals in the network are used to estimate, for each node, the speech component in its reference microphone and loudspeaker signal. For each node then a prediction error method based AFC stage is applied using these estimates to estimate the desired signal. Closed-loop simulations show that the proposed centralized algorithm outperforms a node working in isolation, in terms of the signal-to-interference-plus-noise ratio improvement ($\Delta\mathrm{SINR}$) and short-time objective intelligibility (STOI) metric. In particular, the performance is improved when for instance one of the nodes has a lower input signal-to-noise ratio (iSNR) than the other nodes. In addition, it is also shown that a single-channel local adaptive filter per node is sufficient in the AFC stage, instead of a K-channel adaptive filter. Based on this and on the availability of a distributed algorithm for the NR stage, the presented algorithm is viewed as a stepping stone towards a fully distributed NR and AFC algorithm.

***Index Terms*—** Combined acoustic feedback cancellation and noise reduction, multichannel Wiener filter, prediction error method.

1. INTRODUCTION

Acoustic environments where multiple microphones and loudspeakers are interacting with each other can pose significant challenges to audio signal processing algorithms. Background noise, reverberation and acoustic feedback/echo can corrupt microphone signals and affect the performance of devices such as hearing aids, public address (PA) systems, in-car communication and teleconferencing systems. In particular, acoustic feedback occurs whenever a signal is being recorded, amplified and played back within the same acoustic environment and there is acoustic coupling between a microphone and a loudspeaker. In this paper this acoustic feedback problem is tackled in combination with the background noise problem.

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of the 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 the European Research Council under the European Union's Horizon 2020 Research and Innovation Program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. The scientific responsibility is assumed by its authors.

Existing solutions to combined acoustic feedback cancellation (AFC) and noise reduction (NR) mainly cover single-channel scenarios [1] and hearing aid applications [2, 3]. In [1] the prediction error method (PEM) based adaptive filtering with row operations (PEM-AFROW) algorithm is used in combination with an NR stage based on a minimum mean squared error short-time log spectral amplitude (MMSE-LSA) estimation, for a single-channel scenario. In [2] multiple schemes are presented for combined AFC and NR using a generalised sidelobe canceller (GSC) for the NR stage and a PEM-based AFC stage. In [4] active feedback suppression for one microphone in a hearing aid is proposed using multiple loudspeakers, without considering the presence of noise in the microphone signal. A real-time implementation using a smartphone-based hearing aid is presented in [5] where combined NR and feedback suppression is performed using spectral subtraction. In [6] a blind source separation algorithm is used in combination with an AFC algorithm for a binaural hearing aid.

In [7] the authors have proposed a rank-2 multi-channel Wiener filter (MWF) and single-channel PEM-AFROW algorithm to solve the combined NR and AFC problem. In this paper, building on the results of [7], a centralized cascade multi-channel NR and AFC algorithm is presented in the context of a wireless acoustic sensor and actuator network (WASAN). In a WASAN where nodes may have multiple microphones and loudspeakers, acoustic coupling may occur between any microphone-loudspeaker pair in the network and furthermore each node has its own so-called forward path. Therefore, it can be challenging to avoid acoustic howling in such scenarios. The novelty of the paper is to present an algorithm that offers combined NR and AFC in a WASAN, i.e., with multiple microphones and loudspeakers involved and one forward path in each node. In the considered scenarios there is only one desired source signal, which is common to all the nodes. The algorithm consists of a multi-channel Wiener filter (MWF) based NR stage, where M microphone and L loudspeaker signals in the network are used to estimate, for each node, the speech component in its reference microphone and loudspeaker signal. For each node then a prediction error method based AFC stage is applied using these estimates to estimate the desired signal. It is then also shown by means of simulations, using a two-node WASAN, that only a single-channel local adaptive filter is needed in the AFC stage, instead of K-channel adaptive filter per node (K being the number of nodes in the network). Based on this and on the availability of a distributed algorithm for the NR stage, the presented algorithm is viewed as a stepping stone towards a fully distributed NR and AFC algorithm.

The paper is organized as follows: Section 2 describes the signal model. Section 3 presents the NR and AFC stages of the proposed

algorithm, which is also depicted in Figure 1a. Section 4 presents the simulation results and Section 5 concludes the paper.

2. SIGNAL MODEL

Consider a fully connected WASAN with K nodes where node $k \in \mathcal{K} = \{1, \ldots, K\}$ contains m_k microphones and l_k loudspeakers. The aim of node k is to record a node-independent desired speech signal $s(t)$, amplify it and play it back in its local loudspeakers. The case when $l_k = 1, \forall k$ will be considered, without loss of generality. The loudspeaker signal in node k is denoted by $u_k(t)$ and the mth microphone signal, with $m = 1, \ldots, m_k$, is modeled as

$$x_k^{(m)}(t) = H_k^{(m)}(q,t)s(t) + F_{kk}^{(m)}(q,t)u_{k|s}(t) + \sum_{j\neq k} F_{jk}^{(m)}(q,t)u_{j|s}(t) + n_k^{(m)}(t) \quad (1)$$

where $H_k^{(m)}(q,t)$ is the transfer function from the desired speech source to the mth microphone in node k, $F_{kk}^{(m)}(q,t)$ is the transfer function from the loudspeaker in node k to the mth microphone in node k and $F_{jk}^{(m)}(q,t)$ is the transfer function from the loudspeaker in node j to the mth microphone in node k. The loudspeaker signal and the desired speech component in the loudspeaker signal in node k are denoted by $u_k(t)$ and $u_{k|s}(t)$, respectively. The noise signal in the mth microphone in node k is denoted by $n_k^{(m)}(t)$ and includes the direct noise component in the microphone as well as the noise components in the loudspeaker signals that are coupled into the microphone. The discrete time index is represented by t and q^{-1} is the delay operator, i.e., $q^{-p}u_k(t) = u_k(t-p)$. The loudspeaker signal can be expressed as

$$u_k(t) = \sum_{m=1}^{m_k} G_k^{(m)}(q,t)x_k^{(m)}(t) \quad (2)$$

where $G_k^{(m)}(q,t)$ is the so-called forward path transfer function for the mth microphone in node k. The presence of the forward path creates a closed-loop system which introduces signal correlation between the loudspeaker and microphone signals. It is assumed that the desired speech signal can be modeled as

$$s(t) = A(q,t)e(t) \quad (3)$$

where $A(q,t)$ defines an autoregressive (AR) process and is excited by the white noise signal $e(t)$. The node-specific task in node k is to estimate the desired speech signal without the feedback and noise components, as observed at its chosen reference microphone ($m = r$), i.e.,

$$d_k(t) = H_k^{(r)}(q,t)s(t) \quad (4)$$

where $H_k^{(r)}(q,t)$ is the transfer function from the desired speech source to the reference microphone in node k.

In the short-time Fourier transform (STFT) domain, where the time-frame and frequency-bin indices are l and κ, respectively, an $n_k \times 1$ multi-channel signal vector ($n_k = m_k + 1$), consisting of loudspeaker and microphone signals, can be expressed as

$$\mathbf{y}_k(\kappa,l) = \begin{bmatrix} 0 \\ \mathbf{h}_k(\kappa,l) \end{bmatrix} s(\kappa,l) + \mathbf{z}_k(\kappa,l) + \mathbf{n}_k(\kappa,l) \quad (5)$$

$$= \begin{bmatrix} 0 \\ \mathbf{s}_k(\kappa,l) \end{bmatrix} + \mathbf{z}_k(\kappa,l) + \mathbf{n}_k(\kappa,l) \quad (6)$$

where $s(\kappa,l)$ and $\mathbf{n}_k(\kappa,l)$ are the STFT representation of the desired speech signal and the noise in the microphone and loudspeaker signals in node k, respectively. The acoustic transfer functions from the desired speech source position to the local microphones are collected in $\mathbf{h}_k(\kappa,l)$. The component $\mathbf{z}_k(\kappa,l)$ is defined as

$$\mathbf{z}_k(\kappa,l) = \begin{bmatrix} 1 \\ \mathbf{f}_{kk}(\kappa,l) \end{bmatrix} u_{k|s}(\kappa,l) + \sum_{j\neq k} \begin{bmatrix} 0 \\ \mathbf{f}_{jk}(\kappa,l) \end{bmatrix} u_{j|s}(\kappa,l) \quad (7)$$

where $u_{k|s}(\kappa,l)$, $u_{j|s}(\kappa,l)$ and $u_k(\kappa,l)$ are the STFT representations of $u_{k|s}(t)$, $u_{j|s}(t)$ and $u_k(t)$, respectively. It is noted that $\mathbf{n}_k(\kappa,l)$, similar to $n_k^{(m)}(t)$ in (1), includes the direct noise component in the microphones, the noise components in the loudspeaker signals that are coupled into the microphones and the noise component in the loudspeaker signals. The STFT representations of the feedback paths from the loudspeakers in node k and in node j to the microphones in node k are respectively denoted by $\mathbf{f}_{kk}(\kappa,l)$ and $\mathbf{f}_{jk}(\kappa,l)$. For brevity l and κ will be mostly omitted in the following. Finally the signal vector $\mathbf{y}_k$ can be expressed as

$$\mathbf{y}_k = \begin{bmatrix} u_k^* & \mathbf{x}_k^H \end{bmatrix}^H \quad (8)$$

$$= \tilde{\mathbf{s}}_k + \tilde{\mathbf{f}}_{kk}u_{k|s} + \sum_{j\neq k}^{K} \tilde{\mathbf{f}}_{jk}u_{j|s} + \mathbf{n}_k = \tilde{\mathbf{s}}_k + \mathbf{z}_k + \mathbf{n}_k \quad (9)$$

with $\tilde{\mathbf{s}}_k = \begin{bmatrix} 0 & \mathbf{s}_k^H \end{bmatrix}^H$, $\tilde{\mathbf{h}}_k = \begin{bmatrix} 0 & \mathbf{h}_k^H \end{bmatrix}^H$, $\tilde{\mathbf{f}}_{kk} = \begin{bmatrix} 1 & \mathbf{f}_{kk}^H \end{bmatrix}^H$ and $\tilde{\mathbf{f}}_{jk} = \begin{bmatrix} 0 & \mathbf{f}_{jk}^H \end{bmatrix}^H$.

The N-dimensional vectors, with $N = \sum_{k=1}^{K} n_k$, $\mathbf{y}, \mathbf{s}, \mathbf{z}$ and $\mathbf{n}$ are the stacked versions of $\mathbf{y}_k, \tilde{\mathbf{s}}_k, \mathbf{z}_k$ and $\mathbf{n}_k$ $\forall k$, respectively, such that the network-wide signal vector $\mathbf{y}$ can be written as follows

$$\mathbf{y} = \mathbf{s} + \mathbf{z} + \mathbf{n}. \quad (10)$$

3. CENTRALIZED CASCADE NR AND AFC ALGORITHM

The centralized cascade NR and AFC algorithm of [7] which is adopted here, consists of an N-channel NR stage performed first and followed by a K-channel AFC stage. In the NR stage, a rank-2 MWF is used and two signal estimates per node are obtained, one for the desired speech component in the local reference microphone and the other for the desired speech component in the local loudspeaker. In the AFC stage, a K-channel adaptive filter per node is used to cancel out the feedback component in the noise-reduced reference signal, due to the presence of K loudspeakers in the scenario, and hence multiple feedback paths.

3.1. NR stage

The NR stage aims to estimate the desired speech component in a reference microphone, hence an N-dimensional rank-2 MWF will be used, which requires estimates of the correlation matrices of the signal vector $\mathbf{y}$ and the desired speech component $\mathbf{s}$. In the STFT domain, the N-dimensional correlation matrix of the signal vector $\mathbf{y}$ is

$$\bar{\mathbf{R}}_{\mathbf{yy}} = E\{\mathbf{y}\mathbf{y}^H\} = \bar{\mathbf{R}}_{\mathbf{yy|ss}} + \bar{\mathbf{R}}_{\mathbf{yy|nn}} \quad (11)$$

where $\bar{\mathbf{R}}_{\mathbf{yy|ss}}$ and $\bar{\mathbf{R}}_{\mathbf{yy|nn}}$ denote the speech and noise correlation matrix, respectively, and are defined as

$$\bar{\mathbf{R}}_{\mathbf{yy|nn}} = E\{\mathbf{n}\mathbf{n}^H\}, \quad (12)$$

$$\bar{\mathbf{R}}_{\mathbf{yy|ss}} = E\{(\mathbf{s}+\mathbf{z})(\mathbf{s}+\mathbf{z})^H\}. \quad (13)$$

For a WASAN with 1 loudspeaker in each node, the speech correlation matrix can be defined as

$$\mathbf{P} = \begin{bmatrix} \tilde{\mathbf{f}}_{11} & \cdots & \tilde{\mathbf{f}}_{K1} & \tilde{\mathbf{h}}_1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{\mathbf{f}}_{1K} & \cdots & \tilde{\mathbf{f}}_{KK} & \tilde{\mathbf{h}}_K \end{bmatrix}, \quad (14)$$

$$\mathbf{V} = \begin{bmatrix} u_{1|s}^* & \cdots & u_{K|s}^* & s^* \end{bmatrix}^H, \quad (15)$$

$$\bar{\mathbf{R}}_{\mathbf{yy|ss}} = E\{\mathbf{P}\mathbf{V}\mathbf{V}^H\mathbf{P}^H\}. \tag{16}$$

The minimization of the mean squared error (MSE) between the desired signal and the filtered microphone and loudspeaker signals defines the optimal filter [8]

$$\bar{\mathbf{W}}_k = \min_{\mathbf{W}_k} E\left\{\|\mathbf{d}_k - \mathbf{W}_k^H\mathbf{y}\|^2\right\}. \tag{17}$$

with $\mathbf{d}_k = \begin{bmatrix} u_{k|s}^* & x_{k|s}^{(r)*} \end{bmatrix}^H$, where $u_{k|s}$ represents the speech component in the loudspeaker signal in node k and $x_{k|s}^{(r)}$ represents the total contribution of s together with $u_{1|s}, u_{1|s}, \ldots, u_{K|s}$ in the reference microphone signal in node k. The solution to (17) is the MWF, given by [8, 9]

$$\bar{\mathbf{W}}_k = \bar{\mathbf{R}}_{\mathbf{yy}}^{-1}\bar{\mathbf{R}}_{\mathbf{yy|ss}}[\mathbf{e}_k^{(r)}|\mathbf{e}_k^{(r+1)}] \tag{18}$$

where $\mathbf{e}_k^{(r)}$ and $\mathbf{e}_k^{(r+1)}$ are vectors which select the columns of the matrix that correspond to the loudspeaker and reference microphone in node k, respectively . This expression is obtained based on the assumption that s and $\mathbf{n}$ are uncorrelated. Due to the unavailability of the speech correlation matrix, given that the steering vectors are unknown, a voice activity detector (VAD) is used to estimate *speech-plus-noise* and *noise-only* correlation matrix. In the former, the desired speech and noise are active, in the latter only the noise is active. A recursive update of this matrix pencil is used based on the VAD, i.e.,

$$\begin{aligned} &\text{if VAD}(\kappa, l) = 1: \\ &\hat{\mathbf{R}}_{\mathbf{yy}}(\kappa, l) = \beta\hat{\mathbf{R}}_{\mathbf{yy}}(\kappa, l-1) + (1-\beta)\mathbf{y}(\kappa, l)\mathbf{y}^H(\kappa, l) \end{aligned} \tag{19}$$

$$\begin{aligned} &\text{if VAD}(\kappa, l) = 0: \\ &\hat{\mathbf{R}}_{\mathbf{yy|nn}}(\kappa, l) = \beta\hat{\mathbf{R}}_{\mathbf{yy|nn}}(\kappa, l-1) + (1-\beta)\mathbf{y}(\kappa, l)\mathbf{y}^H(\kappa, l) \end{aligned} \tag{20}$$

where $\hat{\mathbf{R}}_{\mathbf{yy}}(\kappa, l)$ and $\hat{\mathbf{R}}_{\mathbf{yy|nn}}(\kappa, l)$ represent estimates of $\bar{\mathbf{R}}_{\mathbf{yy}}(\kappa, l)$ and $\bar{\mathbf{R}}_{\mathbf{yy|nn}}(\kappa, l)$ at time frame l and frequency bin index κ. The speech correlation matrix can then be estimated via [9]

$$\begin{aligned} \hat{\mathbf{R}}_{\mathbf{yy|ss}} = \min_{\mathbf{R}_{\mathbf{yy|ss}}} &\left\|\hat{\mathbf{R}}_{\mathbf{yy|nn}}^{-1/2} \times \right. \\ &\left.\left(\hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{R}}_{\mathbf{yy|nn}} - \mathbf{R}_{\mathbf{yy|ss}}\right)\hat{\mathbf{R}}_{\mathbf{yy|nn}}^{-H/2}\right\|_F^2 \end{aligned} \tag{21}$$

$$\begin{aligned} \text{s.t.} \quad &\text{rank}(\mathbf{R}_{\mathbf{yy|ss}}) = 2, \\ &\mathbf{R}_{\mathbf{yy|ss}} \succeq 0 \end{aligned} \tag{22}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Spatial pre-whitening is applied in (21) by pre- and post- multiplying by $\hat{\mathbf{R}}_{\mathbf{yy|nn}}^{-1/2}$ and $\hat{\mathbf{R}}_{\mathbf{yy|nn}}^{-H/2}$, respectively. The solution to (21) is based on a generalized eigenvalue decomposition (GEVD) of the $(N \times N)$ matrix pencil $\{\hat{\mathbf{R}}_{\mathbf{yy}}, \hat{\mathbf{R}}_{\mathbf{yy|nn}}\}$ [9, 10]

$$\hat{\mathbf{R}}_{\mathbf{yy}} = \hat{\mathbf{Q}}\hat{\mathbf{\Sigma}}_{\mathbf{yy}}\hat{\mathbf{Q}}^H, \tag{23}$$

$$\hat{\mathbf{R}}_{\mathbf{yy|nn}} = \hat{\mathbf{Q}}\hat{\mathbf{\Sigma}}_{\mathbf{yy|nn}}\hat{\mathbf{Q}}^H \tag{24}$$

where $\hat{\mathbf{\Sigma}}_{\mathbf{yy}}$ and $\hat{\mathbf{\Sigma}}_{\mathbf{yy|nn}}$ are diagonal matrices and $\hat{\mathbf{Q}}$ is an invertible matrix. The rank-2 speech correlation matrix estimate $\hat{\mathbf{R}}_{\mathbf{yy|ss}}$ is then

$$\hat{\mathbf{R}}_{\mathbf{yy|ss}} = \hat{\mathbf{Q}}\text{diag}\{\hat{\sigma}_{yy,1} - \hat{\sigma}_{yy|nn,1}, \hat{\sigma}_{yy,2} - \hat{\sigma}_{yy|nn,2}, 0, \ldots, 0\}\hat{\mathbf{Q}}^H \tag{25}$$

where $\hat{\sigma}_{yy,i}$ and $\hat{\sigma}_{yy|nn,i}$ are the ith diagonal element of $\hat{\mathbf{\Sigma}}_{\mathbf{yy}}$ and $\hat{\mathbf{\Sigma}}_{\mathbf{yy|nn}}$, respectively, which correspond to the ith largest ratio $\hat{\sigma}_{yy,i}/\hat{\sigma}_{yy|nn,i}$. Using (23) and (25) in (18), the rank-2 MWF estimate $\hat{\mathbf{W}}_k$ is expressed as

$$\begin{aligned} \hat{\mathbf{W}}_k = \hat{\mathbf{Q}}^{-H}\text{diag}\Big\{1 - \frac{\hat{\sigma}_{yy|nn,1}}{\hat{\sigma}_{yy,1}}, 1 - \frac{\hat{\sigma}_{yy|nn,2}}{\hat{\sigma}_{yy,2}}, \\ 0, \ldots, 0\Big\}\hat{\mathbf{Q}}^H[\mathbf{e}_k^{(r)}|\mathbf{e}_k^{(r+1)}]. \end{aligned} \tag{26}$$

The desired signal estimates are then obtained as $\hat{u}_{k|s} = \mathbf{e}_k^{(r)H}\hat{\mathbf{W}}_k\mathbf{y}$ and $\hat{x}_{k|s}^{(r)} = \mathbf{e}_k^{(r+1)H}\hat{\mathbf{W}}_k\mathbf{y}$.

3.2. AFC stage

The NR stage provides an estimate for

$$x_{k|s}^{(r)}(t) = H_k^{(r)}(q, t)s(t) + \sum_{j=1}^{K} F_{jk}^{(r)}(q, t)u_{j|s}(t) \tag{27}$$

from which the AFC stage will now estimate $H_k^{(r)}(q, t)s(t)$. A K-channel so-called PEM-AFROW algorithm is used. This algorithm was initially developed in [11] for a single-channel scenario and it provides estimates of both the feedback path and the speech source signal model. The PEM-AFROW algorithm used here is the frequency-domain version presented in [12] (the reader is referred to [12] for a detailed explanation of the AFC algorithm). The algorithm uses an overlap-save (OLS) filterbank to compute convolutions in the frequency domain, which requires a rectangular window [13]. The input signals to the AFC algorithm in node k are $\hat{u}_{1|s}, \ldots, \hat{u}_{K|s}$ and $\hat{x}_{k|s}^{(r)}(t)$.

It is also proposed here to only use a single-channel local adaptive filter per node, instead of the K-channel AFC filter, motivated by the high correlation between all loudspeaker signals. This is equivalent to assuming a linear dependency between all the loudspeaker signals [6], which in a scenario with only one desired source is reasonable. It is noted, however, that in an scenario where the nodes are not interested in the same desired source signal or the forward path transfer functions in each node are non-linear, this assumption will not be a valid one. The assumption will be verified in Section 4. Figure 1a shows a block diagram of the complete closed-loop system including the proposed centralized cascade NR and AFC algorithm for a two-node scenario.

4. SIMULATION RESULTS

Closed-loop simulation results are presented in this section using a scenario consisting of a two-node WASAN where node 1 has a four-microphone linear array and one loudspeaker, and node 2 has a two-microphone linear array and one loudspeaker. Artificial impulse responses were generated using the randomized image method in [14] and the near-end signal was speech. No reverberation was used and a white noise component was added to the microphones. The forward path gains were the same for each node, and were specified by the gain profile shown in Figure 1b. The gain profile was chosen such that the *noise-only* and *speech-plus-noise* correlation matrices could be updated while the system is stable, then it is gradually increased to test the algorithms. The loudspeaker signal in all scenarios was obtained by using the feedback-compensated signal, multiplied and delayed by the forward path gain and delay respectively. The impulse response length was $1024\,\text{samples}$. The maximum stable gain (MSG), K_{MSG}, was found experimentally, by running simulations without any algorithm and with different gain values until the system became unstable. The forward path delay depends on the window size used for both the weighted overlap-add (WOLA) and OLS procedures. Here, it was set to $\frac{3R}{2}$, with $R = \frac{M}{2}$. An M-samples long root-squared Hann window was used for the analysis and synthesis

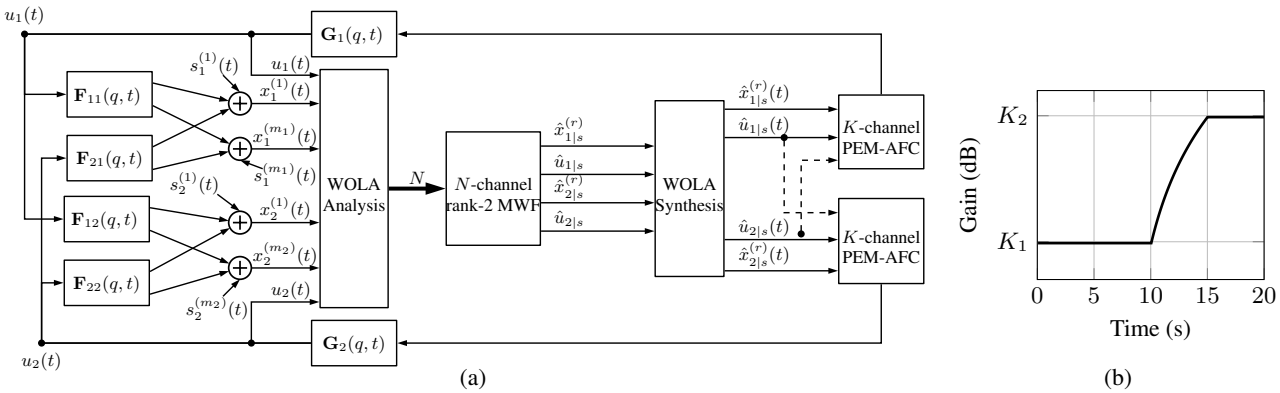

Fig. 1: (a) Closed-loop system including the proposed centralized cascade NR and AFC algorithm for a two-node scenario. Here the symbol $s_k^{(m)}$ represents the desired speech component at microphone m in node k. (b) Forward path gain profile for node 1 and 2, with $K_1 = K_{\text{MSG}} - 10$ and $K_2 = K_{\text{MSG}}$.

filterbank in WOLA for the NR stage and an M-samples long rectangular window was used in the OLS filterbank for the AFC stage. The scenario was simulated using a fixed input signal-to-noise ratio (SNR) (iSNR) of 10 dB for node 1 and an iSNR in the range $\{-15, 20\}$ dB for node 2. An oracle VAD was used, but practical implementations can be found in [15, 16]. Two configurations of the algorithm are investigated, namely centralized and isolated. The proposed centralized configuration follows the algorithm described in Figure 1a. In the isolated algorithm each node performs the algorithm in [7], i.e., only local microphone and loudspeaker signals are used hence there is no cooperation with other nodes. The performance of the algorithm is assessed using objective and perceptual measures, namely the signal-to-interference-plus-noise ratio (SINR) and short-time objective intelligibility (STOI). STOI is a correlation-based speech intelligibility measure that works on the temporal envelopes of short speech frames [17, 18]. In this paper, we used a MATLAB implementation of the STOI from [17]. It was chosen based on the results presented in [18] where objective and subjective metrics were assessed for AFC algorithms. The SINR is defined as

$$\text{SINR}_k(l) = 10\log_{10}\left(\frac{\sum_{t=0}^{R-1} v_k^2(t-lR)}{\sum_{t=0}^{R-1} i_k^2(t-lR)}\right) \quad \text{dB}. \tag{28}$$

where the input SINR (iSINR) can be computed with $v_k(t) = H_k^{(m)}(q,t)s(t)$ and $i_k(t) = \sum_{j=1}^{K} F_{jk}^{(r)}(q,t)u_j(t) + n_k^{(r)}(t)$, R is the window size used to compute the SINR. Similarly, the output SINR (oSINR) is computed using the estimates of the quantities of the iSINR, which means each component needs to be filtered separately with the filter estimates in both the NR and AFC stages. In this paper, the average SINR improvement is used as a metric, which is defined as

$$\Delta\text{SINR}_k = \frac{1}{L}\sum_{l=1}^{L} \text{oSINR}_k(l) - \text{iSINR}_k(l) \quad \text{dB}. \tag{29}$$

where L is the total number of blocks where the desired speech source is active. In Figure 2 it can be observed that the proposed centralized algorithm outperforms the isolated configuration (the legends Centralized 1-AF and Centralized 2-AF refer to a centralized configuration with 1 and 2 adaptive filters per node in the AFC stage, respectively). The cooperation clearly improves the ΔSINR and STOI metrics at node 2 when its iSNR is worse than the iSNR at node 1. The performance in node 1 is almost the same between the centralized and isolated algorithms for both metrics. This is due to the iSNR at node 1, which is already good enough to perform the

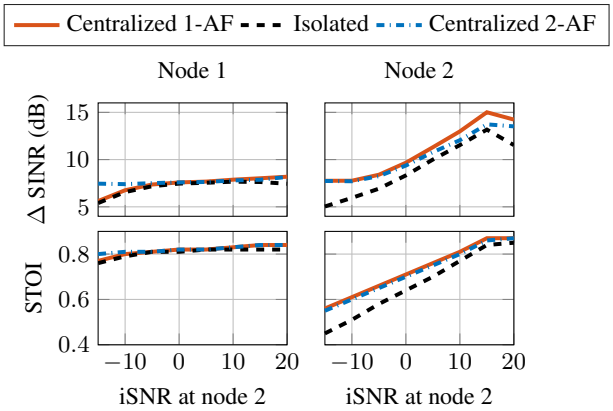

Fig. 2: ΔSINR and STOI for node 1 and 2 for the centralized and isolated algorithms.

NR and AFC algorithm using only its local signals. It is also noted that the performance of Centralized 1-AF and Centralized 2-AF is not so different, except in node 1 when the iSNR at node 2 is -10 and -15 dB. Even though the desired signal estimate after the NR stage at the reference microphone in a node consists of local and non-local feedback components (cfr. (27)), which suggest multiple adaptive filters should be used, due to the high correlation between the loudspeaker signals, the non-local feedback components can be neglected and only one adaptive filter per node is needed to cancel the feedback components at the reference microphone [6]. It also should be noted that the non-local feedback components will be filtered by the NR stage, which will see it as an extra interference source.

5. CONCLUSION

A centralized cascade multi-channel NR and AFC algorithm has been presented. A scenario in a WASAN with two nodes has been used to compare the performance of the centralized and the isolated algorithm. It is shown by simulations that the centralized algorithm outperforms the isolated one in terms of ΔSINR and STOI. It was also shown that the node with the lowest iSNR benefits more from the cooperation than a node with a high iSNR. Simulations also showed that the centralized algorithm can achieve almost the same performance with only a one-channel AFC filter per node instead of a K-channel filter, which is a step towards developing a distributed algorithm where network-wide the NR stage is performed in a distributed fashion [19] with a post-processing step, i.e., the AFC stage.

6. REFERENCES

[1] A. Bastari, S. Squartini, and F. Piazza, "Joint acoustic feedback cancellation and noise reduction within the prediction error method framework," in *2008 Hands-Free Speech Communication and Microphone Arrays*, 2008, pp. 228–231.

[2] A. Spriet, G. Rombouts, M. Moonen, and J. Wouters, "Combined feedback and noise suppression in hearing aids," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 6, pp. 1777–1790, 2007.

[3] H. Schepker, S. E. Nordholm, L. T. T. Tran, and S. Doclo, "Null-steering beamformer-based feedback cancellation for multi-microphone hearing aids with incoming signal preservation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 4, pp. 679–691, 2019.

[4] H. Schepker and S. Doclo, "Active feedback suppression for hearing devices exploiting multiple loudspeakers," in *Proc. 2015 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '19)*, 2019, pp. 60–64.

[5] M. Vashkevich, E. Azarov, N. Petrovsky, D. Likhachov, and A. Petrovsky, "Real-time implementation of hearing aid with combined noise and acoustic feedback reduction based on smartphone," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6570–6571.

[6] A. Lombard, K. Reindl, and W. Kellermann, "Combination of adaptive feedback cancellation and binaural adaptive filtering in hearing aids," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–15, 2009.

[7] S. Ruiz, T. van Waterschoot, and M. Moonen, "Cascade multichannel noise reduction and acoustic feedback cancellation," in *Proc. 2022 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '22)*, 2022, pp. 676–680.

[8] 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.

[9] 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, 2014.

[10] F. Jabloun and B. Champagne, "Signal subspace techniques for speech enhancement," in *Speech Enhancement*, pp. 135–159. Springer, 2005.

[11] A. Spriet, M. Moonen, and I. Proudler, "Feedback cancellation in hearing aids: an unbiased modelling approach," in *Proc. 11th European Signal Process. Conf. (EUSIPCO '02)*, 2002, pp. 1–4.

[12] G. Bernardi, T. van Waterschoot, J. Wouters, and M. Moonen, "An all-frequency-domain adaptive filter with PEM-based decorrelation for acoustic feedback control," in *Proc. 2015 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '15)*. IEEE, 2015, pp. 1–5.

[13] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, 1992.

[14] 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 Lang. Process.*, vol. 23, no. 4, pp. 774–786, 2015.

[15] V. Berisha, H. Kwon, and A. Spanias, "Real-time implementation of a distributed voice activity detector," in *Proc. 4th IEEE Workshop Sensor Array Multichannel Process.*, pp. 659–662.

[16] Yingke Zhao, Jesper Kjær Nielsen, Jingdong Chen, and Mads Græsbøll Christensen, "Model-based distributed node clustering and multi-speaker speech presence probability estimation in wireless acoustic sensor networks," *J. Acoust. Soc. Amer.*, vol. 147, no. 6, pp. 4189–4201, 2020.

[17] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proc. 2010 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '10)*, 2010, pp. 4214–4217.

[18] G. Bernardi, T. van Waterschoot, J. Wouters, and M. Moonen, "Subjective and objective sound-quality evaluation of adaptive feedback cancellation algorithms," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 5, pp. 1010–1024, 2018.

[19] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks—part i: Sequential node updating," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5277–5291, 2010.