ANALYSIS OF RATE CONSTRAINTS FOR MWF-BASED NOISE REDUCTION IN ACOUSTIC SENSOR NETWORKS

Toby Christian Lawin-Ore, Simon Doclo

University of Oldenburg, Institute of Physics, Signal Processing Group, Oldenburg, Germany

{toby.chris.lawin.ore,simon.doclo}@uni-oldenburg.de

ABSTRACT

In an acoustic sensor network, consisting of spatially distributed microphone nodes, a significant noise reduction can be achieved using the centralized multi-channel Wiener filter (MWF), requiring all available microphone signals in the entire network. However the limited bandwidth of the communication link typically does not allow to transmit all microphone signals between the different nodes. Recently, a distributed node-specific MWF-based noise reduction scheme has been presented, where each node only transmits a filtered combination of its microphone signals. In this paper, the performance gain of the centralized MWF and the distributed node-specific MWF-based scheme are analyzed as a function of the available bandwidth of the communication link.

Index Terms— Acoustic sensor networks, distributed estimation, rate distortion theory

1. INTRODUCTION

Due to current advances in hardware technology it is nowadays possible to produce small microphone devices capable to communicate wirelessly over a short range. By spatially distributing these microphones one can build a so-called *acoustic sensor network* (ASN) with several microphones located at distinct places, e.g. at locations where it is impossible or undesirable to place wired microphones. When distributed over a large area, the acoustic sensor array network is able to collect more information than a microphone array located at a single position. For example, in a binaural hearing aid application, microphone arrays located on different hearing aids (or even other devices) are able to exchange information with each other using a wireless link in order to improve speech intelligibility in noisy environments [1]-[6].

For a binaural link with limited capacity, a theoretically optimal (in an information-theoretic sense) transmission scheme has been presented in [4], however requiring knowledge about the joint statistics of the signals at both hearing aids, which is typically not available in practice. In [5] the relation between performance gain and link capacity has been analyzed for several suboptimal (but practical) schemes, such as transmitting one microphone signal or transmitting an estimate of the desired signal obtained at the transmitting device. In [7] a similar relation between performance gain and link capacity has been presented for the iterative distributed MWF scheme [3], which (similarly as in [4] and [5]) has been designed for an ASN with only 2 nodes.

Recently, a more general distributed adaptive node-specific signal estimation (DANSE) algorithm has been introduced for signal estimation in an acoustic sensor network with more than two nodes [6]. In [6][8] the benefit of this noise reduction scheme, where the different nodes in the ASN can exchange information to improve their performance, has been shown, however assuming an infinite bandwidth for transmitting the microphone signals. In this paper the effect on the performance due to the limited capacity of the wireless link will be analyzed.

This paper is organized as follows. Section 2 presents the signal model and the sensor network configuration. In section 3 the centralized MWF is reviewed. Section 4 introduces the DANSE algorithm. In section 5 we address the rate distortion theory, which enables to link the transmission rate to the mean-squared-error (MSE). Finally, in section 6 simulation results are presented for an example with 3 microphone array nodes, where it is shown that the distributed MWF-based algorithm performs, under specific rate constraints, better than the centralized MWF.

2. SIGNAL MODEL AND SENSOR NETWORK CONFIGURATION

Consider the acoustic sensor network with J nodes depicted in Figure 1, where each node n has a microphone array consisting of M_n microphones, with the total number of microphones $M = \sum_{n=1}^{J} M_n$. The m-th microphone signal $Y_{n,m}(\omega)$ of the n-th node can be described in the frequency domain as

$$Y_{n,m}(\omega) = X_{n,m}(\omega) + V_{n,m}(\omega), \quad m = 0 \ldots M_n - 1, \quad (1)$$

where $X_{n,m}(\omega)$ represents the speech component and $V_{n,m}(\omega)$ represents the noise component. For conciseness we will omit the frequency-domain variable ω in the remainder of the paper. We define the M_n-dimensional stacked vectors $\mathbf{Y}_n$ and the M-dimensional signal vector $\mathbf{Y}$ as

$$\mathbf{Y}_n = \begin{bmatrix} Y_{n,0} \\ \vdots \\ Y_{n,M_n-1} \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_J \end{bmatrix}. \quad (2)$$

The network-wide signal vector can hence be written as $\mathbf{Y} = \mathbf{X} + \mathbf{V}$. In the case of a single desired speech source, the speech signal vector can be written as $\mathbf{X} = \mathbf{A}S$, where the steering vector $\mathbf{A}$ containing the acoustic transfer functions between the speech source and the microphones is defined similarly as $\mathbf{Y}$ and S denotes the speech signal.

The output signal $\hat{X}_n$ of the n-th node is obtained by filtering and summing the node-specific microphone signals and the transmitted signals from the other nodes. Assuming an ideal fully connected network, i.e. the signals transmitted by one node are received at all other nodes without any distortion (e.g. due to coding), the output signals can be written as a linear combination of all microphone signals in the network, i.e.

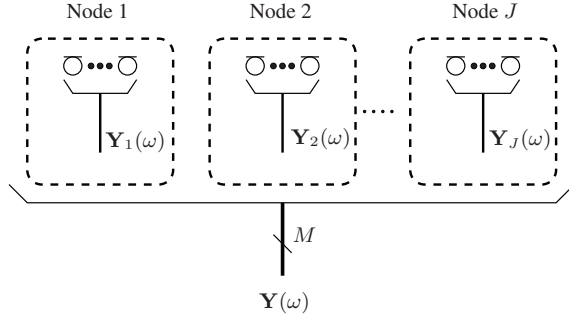$$\hat{X}_n = \mathbf{W}_n^H \mathbf{Y}, \quad n = 1 \ldots J, \quad (3)$$

Fig. 1. Configuration of a sensor network with J sensor nodes

where $\mathbf{W}_n$ is a node-specific M-dimensional filter.

3. CENTRALIZED MULTICHANNEL WIENER FILTERING

In the following we will consider the problem of estimating at each node n as desired signal the speech component $X_{n,0}$ of the first microphone signal selected to be the reference microphone of node n. The node-specific centralized multi-channel Wiener filter (MWF) produces a minimum-mean-square-error (MMSE) estimate at each node by minimizing the MSE cost function [6]

$$\xi(\mathbf{W}_n) = \mathcal{E}\{|X_{n,0} - \mathbf{W}_n^H \mathbf{Y}|^2\}, \qquad (4)$$

where $\mathcal{E}\{.\}$ denotes the expected value operator. Assuming that the speech and noise components are uncorrelated, the multi-channel Wiener filter at node n is given by

$$\mathbf{W}_n^0 = \mathbf{\Phi}_y^{-1} \mathbf{\Phi}_x \mathbf{e}_n, \qquad (5)$$

with $\mathbf{\Phi}_y = \mathcal{E}\{\mathbf{Y}\mathbf{Y}^H\}$ and $\mathbf{\Phi}_x = \mathcal{E}\{\mathbf{X}\mathbf{X}^H\}$ and $\mathbf{e}_n$ an M-dimensional vector with one element equal to 1 and all other elements equal to 0, which selects the column of $\mathbf{\Phi}_x$ corresponding to the reference microphone at node n. By using the multi-channel Wiener filter in (4) the minimum MSE is than equal to

$$\xi(\mathbf{W}_n^0) = \Phi_{x_{n,0}} - \mathbf{\Phi}_{xx_n}^H \mathbf{\Phi}_y^{-1} \mathbf{\Phi}_{xx_n},$$

where $\mathbf{\Phi}_{xx_n} = \mathbf{\Phi}_x \mathbf{e}_n$ and $\Phi_{x_{n,0}} = \mathcal{E}\{|X_{n,0}|^2\}$.

4. DISTRIBUTED MWF IN A SENSOR NETWORK

To compute the node-specific centralized MWF, each node n needs to transmit its unprocessed microphone signals to all other nodes in the network, which requires a large communication bandwidth, obviously depending on the number of nodes and on the number of microphones at each node. As described in [6][8], the DANSE algorithm aims to compute the network-wide MWF in a distributed fashion, hence reducing the communication bandwidth. Figure 2 shows a schematic illustration of the DANSE scheme.

We will now briefly review the DANSE algorithm. For more details we refer to [6][8]. Redefine the filter $\mathbf{W}_n$ at each node as

$$\mathbf{W}_n = \left[\begin{array}{cccc} \mathbf{W}_{n1}^T \mathbf{W}_{n2}^T \ldots \mathbf{W}_{nJ}^T \end{array} \right]^T, \qquad (6)$$

where $\mathbf{W}_{nk}$, $k = 1 \ldots J$, can be viewed as a partial estimator that is specific to node n. Instead of transmitting all microphone signals,
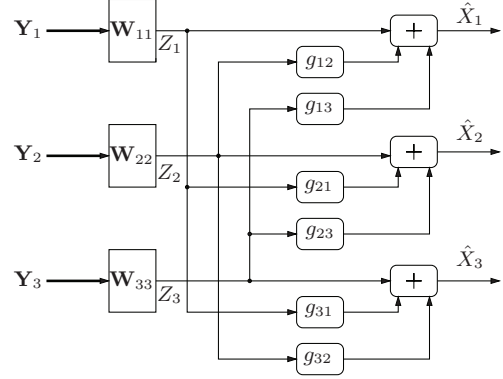

Fig. 2. The DANSE scheme with $J = 3$ sensor nodes

we now consider the case where each node n only transmits the signal $Z_n = \mathbf{W}_{nn}^H \mathbf{Y}_n$. The output signal at each node can then be written as a linear combination of the node-specific signals and the received signals from the other nodes, i.e.

$$\hat{X}_n = \mathbf{W}_{nn}^H \mathbf{Y}_n + \sum_{k=1, k \neq n}^{J} g_{nk}^H Z_k \qquad (7)$$

Define the vectors $\mathbf{Z}_n = \left[Z_1 \cdots Z_J \right]^T$ and $\mathbf{g}_n = \left[g_{n1} \cdots g_{nJ} \right]^T$. The vectors $\mathbf{Z}_{-n}$ and $\mathbf{g}_{n-k}$ are defined as the vectors $\mathbf{Z}_n$ and $\mathbf{g}_n$ with the elements Z_n and g_{nk} respectively omitted. Equation (7) can now be rewritten as

$$\hat{X}_n = \mathbf{W}_{nn}^H \mathbf{Y}_n + \mathbf{g}_{n-n}^H \mathbf{Z}_{-n} = \tilde{\mathbf{W}}_n^H \tilde{\mathbf{Y}}_n,$$

with

$$\tilde{\mathbf{W}}_n = \begin{bmatrix} \mathbf{W}_{nn} \\ \mathbf{g}_{n-n} \end{bmatrix}, \quad \tilde{\mathbf{Y}}_n = \begin{bmatrix} \mathbf{Y}_n \\ \mathbf{Z}_{-n} \end{bmatrix}.$$

The DANSE-algorithm evaluated in this paper iteratively updates the filter coefficients in a round-robin fashion, i.e. at each iteration one specific node updates its parameters. If we denote the filters and the signals in the ith iteration with superscript i, then the iterative procedure of the DANSE algorithm runs as follows:

1. Initialization[1]: $i \leftarrow 0, p \leftarrow 1$
   Initialize $\mathbf{W}_{nn}^0$ and $\mathbf{g}_{n-n}^0$ with random vectors, $\forall n$

2. Each node transmits its partial estimate to the other nodes. Update the parameters of each node according to:

$$\begin{bmatrix} \mathbf{W}_{nn}^{i+1} \\ \mathbf{g}_{n-n}^{i+1} \end{bmatrix} = \begin{cases} (\mathbf{\Phi}_{\tilde{\mathbf{y}}_n}^i)^{-1} \mathbf{\Phi}_{\tilde{\mathbf{x}}_n}^i \tilde{\mathbf{e}}_n & \text{if } n = p \\ \begin{bmatrix} \mathbf{W}_{nn}^i \\ \mathbf{g}_{n-n}^i \end{bmatrix} & \text{if } n \neq p \end{cases} \qquad (8)$$

3. Compute the estimate of the desired signal at each node using the multi-channel Wiener filter $\tilde{\mathbf{W}}_n^{i+1}$.

4. $i \leftarrow i + 1$,
   $p \leftarrow (p \mod J) + 1$

5. return to step 2.

---

[1]Notice that the parameter p indicates which node in the network is allowed to update its parameters.

The $(M_n + J - 1)$-dimensional vector $\tilde{\mathbf{e}}_n$ is defined similarly to $\mathbf{e}_n$, i.e. it selects the reference microphone of node n. Compared to the centralized MWF, each node n now has access to only $M_n + J - 1$ signals instead of M signals, which leads to a reduction of the dimension of the matrices $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_v$.

In the case of a single speech source and in an ideal fully connected network, i.e. the signals transmitted by a node can be received by all other nodes in the network without any coding distortion, it has been shown in [8] that the DANSE algorithm converges for any initialization of its parameters to the centralized MWF for all nodes.

5. CAPACITY OF WIRELESS LINK

Whereas in [6] an ideal communication link (with infinite bandwidth) between the nodes has been assumed, in this paper we will analyze the influence of the available capacity of the link on the performance of the centralized MWF and the DANSE algorithm.

When the transmitted signal Z_n from node n is compressed at rate R (bits per sample), the following rate-distortion relation holds [9]:

$$R(\lambda) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \max\left(0, \log_2 \frac{\Phi_{z_n}(\omega)}{\lambda}\right) d\omega , \quad (9)$$

$$D(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \min\left(\lambda, \Phi_{z_n}(\omega)\right) d\omega , \quad (10)$$

with the parameter λ linking the rate R and the distortion D, and Φ_{z_n} the power spectral density (PSD) of the signal Z_n. Using the forward channel representation [9], depicted in Figure 3, the compressed signal $\tilde{Z}_n$ received at all other nodes can be represented as $\tilde{Z}_n = B Z_n + W$, where B is a bandpass filter with frequency response

$$B = \max\left(0, \frac{\Phi_{z_n} - \lambda}{\Phi_{z_n}}\right) ,$$

and W is additive Gaussian noise with PSD

$$\Phi_W = \max\left(0, \lambda \frac{\Phi_{z_n} - \lambda}{\Phi_{z_n}}\right) .$$

It should be noted that using such a representation in the analysis provides an upper bound on the achievable performance at rate R.
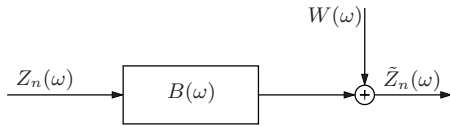

Fig. 3. Forward channel representation

6. EXPERIMENTAL RESULTS

In this section the performance of the centralized MWF and the DANSE algorithm are compared as a function of the capacity of the communication link.

6.1. Setup and performance measures

Simulations have been performed using the simple acoustic scenario depicted in Figure 4. This sensor network with $J = 3$ nodes represents the case of two hearing aid users listening to one speaker,
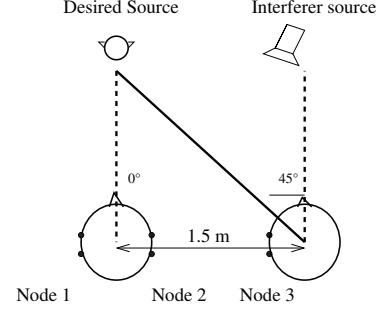

Fig. 4. The scenario of an acoustic sensor network with $J = 3$ nodes

who produces a desired signal. The number of microphones on each hearing aid is $M_n = 2$ and the distance between the microphones on each hearing aid is $1\,\mathrm{cm}$. We consider a scenario with a single speech source S, a single interference I and spatially uncorrelated noise on each microphone, such that the microphone signal vector $\mathbf{Y}$ can be written as

$$\mathbf{Y} = \mathbf{A}_s S + \mathbf{A}_i I + \mathbf{U} , \quad (11)$$

where $\mathbf{A}_s$ and $\mathbf{A}_i$ represent the acoustic transfer functions for the speech source and the interference, respectively, and $\mathbf{U}$ represents spatially uncorrelated noise. Since the speech source, interference and noise are assumed to be uncorrelated, the correlation matrix $\mathbf{\Phi}_y$ is equal to

$$\mathbf{\Phi}_y = \Phi_s \mathbf{A}_s \mathbf{A}_s^H + \Phi_i \mathbf{A}_i \mathbf{A}_i^H + \Phi_u \mathbf{I}_{2M} , \quad (12)$$

with Φ_s, Φ_i and Φ_u the PSDs of the speech source, interference and noise, respectively. All involved PSDs are assumed to be flat in the band [-Ω,Ω], where $\Omega = 2\pi F$ and $F = 8\,\mathrm{kHz}$. The signal-to-interference ratio (SIR) and the signal-to-noise ratio (SNR) are defined as

$$\mathrm{SIR} = 10 \log_{10} \frac{\Phi_s}{\Phi_i} , \quad \mathrm{SNR} = 10 \log_{10} \frac{\Phi_s}{\Phi_u} . \quad (13)$$

The acoustic transfer functions $\mathbf{A}_s$ and $\mathbf{A}_i$ are modeled using the spherical head shadow model in [10] with a radius of $8.75\,\mathrm{cm}$, without taking into account reverberation. Note however that the PSD Φ_{z_n} is non-flat due to the non-flat acoustic transfer functions.

As in [4, 5, 7], the performance gain is defined as the ratio between the MSE at rate 0 and the MSE at rate R, i.e.

$$G(R) = 10 \log_{10} \frac{\xi(0)}{\xi(R)} , \quad (14)$$

which represents the gain in dB due to the availability of the wireless link. $\xi(R)$ denotes the integrated MSE over all considered frequencies obtained when estimating the desired signal at node n when the total available rate in the network is equal to R. We will only consider the performance gain at node 1, but the same analysis holds when considering the other nodes.

6.2. Results

For SIR$= 0\,\mathrm{dB}$ and SNR$= 20\,\mathrm{dB}$, Figures 5 and 6 compare the performance gain of the DANSE algorithm with the centralized MWF. For the centralized MWF the total rate R is equally distributed between the microphone signals, i.e. the m-th microphone signal of
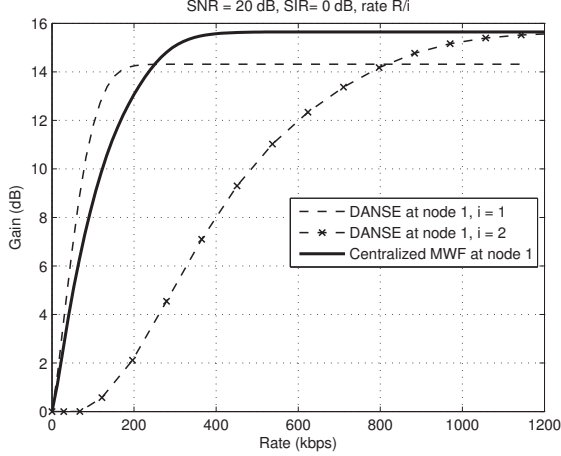
Fig. 5. Performance gain as a function of link capacity for centralized MWF and DANSE for different number of iterations (SIR= 0 dB, SNR= 20 dB, rate R/i).
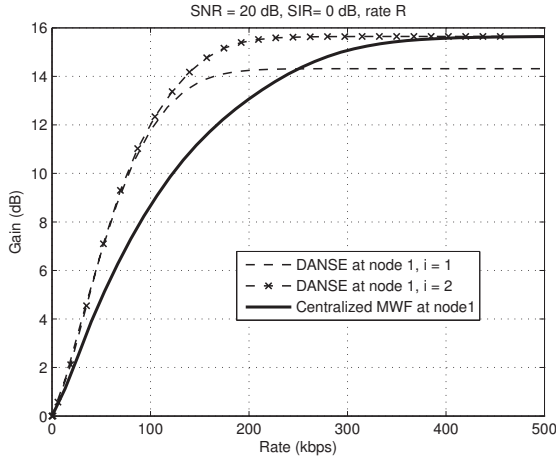

Fig. 6. Performance gain as a function of link capacity for centralized MWF and DANSE for different number of iterations (SIR= 0 dB, SNR= 20 dB, rate R).

the n-th node is compressed at rate $R_{n,m} = R/M$. For the first iteration of the DANSE algorithm ($i = 1$), the filter coefficients $\mathbf{W}_{nn}^1$ are initialized to $\mathbf{W}_{nn}^{1,H} = \begin{bmatrix} 1 & 0 \end{bmatrix}$, i.e. the front microphone signals of the hearing aids are transmitted to the other nodes.

Figure 5 shows the performance gain of the centralized MWF and the DANSE algorithm for a different number of iterations i, where the total rate R is evenly distributed between the iterations, i.e. in each iteration the signals Z_n transmitted from node n to the other nodes are compressed with rate R/i. As can be seen from Figure 5, for low rates the highest performance gain is achieved by transmitting just a single microphone signal (in this case the front signal), corresponding in this figure to $i = 1$. Performing more iterations only leads to an improved performance at very high rates. For example at rate $R = 500$ kbps it is still better to just transmit the single microphone signal instead of performing more iterations. As shown in [6], for a sensor network with ideal communication link (i.e. infinite or large enough bandwidth) the DANSE algorithm converges to the centralized MWF solution, typically requiring only a

small number of iterations.

For many applications it can however be assumed that the signal statistics remain stationary over a (small) number of signal frames, such that the iterations of DANSE can be spread over subsequent frames, instead of performing several iterations on the same frame, as in Figure 5. Figure 6 shows the performance gain of the DANSE algorithm and the centralized MWF, where now in each iteration the signals Z_n are compressed with rate R. This figure shows that for the considered scenario, DANSE converges after $i = 2$ iterations at all rates, moreover achieving the highest performance gain.

7. CONCLUSION

In this paper, the performance gain of the DANSE algorithm compared to the centralized MWF has been analyzed as a function of the capacity of the communication link. It has been shown that the DANSE algorithm converge to the centralized MWF only at high bitrates. When the iterations can be spread over subsequent frames, the DANSE algorithm yields the highest performance gain after only a small number of iterations for all bitrates.

8. REFERENCES

[1] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, *Acoustic beamforming for hearing aid applications*, chapter 9 in "Handbook on Array Processing and Sensor Networks", pp. 269–302, Wiley, 2010.

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

[3] S. Doclo, T. van den Bogaert, J. Wouters, and M. Moonen, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no.1, pp. 38–51, Jan. 2009.

[4] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," *IEEE Trans. Signal Processing*, vol. 57, no. 2, pp. 645–657, Feb. 2009.

[5] S. Srinivasan and A. den Brinker, "Rate-constrained beamforming in binaural hearing aids," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 257197, 2009.

[6] A. Bertrand and M. Moonen, "Robust distributed noise reduction in hearing aids with external acoustic sensor nodes," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 530435, 2009.

[7] S. Doclo, T. C. Lawin-Ore, and T. Rohdenburg, "Rate-constrained binaural MWF-based noise reduction algorithms," *in Proc. ITG Conference Speech Communication*, Bochum, Germany, Oct. 2010.

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

[9] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice Hall, 1971.

[10] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, Nov. 1998.