Reference Microphone Selection for MWF-based Noise Reduction Using Distributed Microphone Arrays

Toby Christian Lawin-Ore, Simon Doclo

University of Oldenburg, Institute of Physics, Signal Processing Group, Oldenburg, Germany
Email: {toby.chris.lawin.ore,simon.doclo}@uni-oldenburg.de
Web: www.sigproc.uni-oldenburg.de

Abstract

Using an acoustic sensor network, consisting of spatially distributed microphones, a significant noise reduction can be achieved with the centralized multi-channel Wiener filter (MWF), which aims to estimate the desired speech component in one of the microphones, referred to as the reference microphone. However, since the distributed microphones are typically placed at different locations, the selection of the reference microphone has a significant impact on the performance of the MWF, largely depending on the position of the desired source with respect to the microphones. In this paper, different optimal and suboptimal reference selection procedures are presented, both broadband and frequency-dependent. Experiment results show that the proposed procedures yield better performance than an arbitrarily selected reference microphone.

1 Introduction

By spatially distributing several microphones, one can build a so-called *acoustic sensor network* (ASN) with microphones located at distinct places, such that more information about the sound field can be acquired than using a single microphone (array) at one position and the probability that a subset of microphones is closer to the desired source(s) is substantially increased. Recently, ASNs have been considered for teleconferencing applications [1], surveillance [2] and for hearing aid applications [3]-[7], where microphone arrays located on different hearing aids (or even other devices) exchange information with each other in order to improve speech intelligibility in noisy environments.

In speech enhancement applications, the muli-channel Wiener filter (MWF) is widely used to reduce noise and thus improve signal quality [6]. The MWF performs noise reduction by estimating the desired signal component in one of the microphones, referred to as the reference microphone. In [8], the theoretical performance measures of the MWF have been analyzed and it has been shown that the theoretical output SNR of the MWF only depends on the noise field and the acoustic transfer functions (ATFs) between the desired source and the microphones. Although, the theoretical output SNR of the MWF is independent of the selection of the reference microphone, experimental results (cf. Table 1) have shown that the estimation of the desired signal component in different reference microphones leads to different output SNRs, depending on the acoustical scenario, i.e., the positions of speech/noise sources and the microphones. This effect can be explained by the fact that for practical implementation of the MWF, the correlation matrices of the speech and noise components are used and estimation errors in these second-order statistics lead to different output SNRs for different reference microphones. For microphone arrays with closely spaced microphones, the impact of the selected reference microphone on the performance is typically small. However, when the microphones are distributed at distinct locations (e.g., in a room), the influence of the reference microphone selection on the performance of the MWF can be quite substantial.

This work was supported by the Research Unit FOR 1732 "Individualized Hearing Acoustics", funded by the German Research Foundation (DFG).

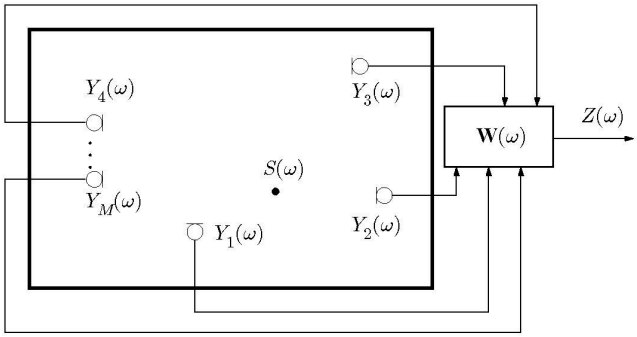

Figure 1: Configuration of a sensor network with M microphones.

In this paper, we introduces different optimal and suboptimal (broadband and frequency-dependent) reference selection procedures, based on output/input SNR, signal energy and source distance and investigate the performance of the MWF using these procedures. Simulation results have shown that the reference selection procedures proposed in this paper perform better than an arbitrarily selected reference microphone. Moreover, the less complex suboptimal procedures show similar performance as the optimal reference selection procedures.

2 Signal model and configuration

Consider the acoustic sensor network with M distributed microphones as depicted in Figure 1. The m-th microphone signal $Y_m(\omega)$ can be written in the frequency-domain as

$$Y_m(\omega) = X_m(\omega) + V_m(\omega) \quad m = 1 \ldots M, \tag{1}$$

where $X_m(\omega)$ represents the speech component and $V_m(\omega)$ the noise component. We define the M-dimensional stacked vector $\mathbf{Y}(\omega)$ as

$$\mathbf{Y}(\omega) = \begin{bmatrix} Y_1(\omega) \\ \vdots \\ Y_M(\omega) \end{bmatrix}, \tag{2}$$

which can be decomposed as $\mathbf{Y}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega)$.

The noise reduced signal is then obtained by filtering and summing the microphone signals, i.e.,

$$Z(\omega) = \mathbf{W}^H(\omega)\mathbf{X}(\omega) + \mathbf{W}^H(\omega)\mathbf{V}(\omega), \tag{3}$$

where $\mathbf{W}(\omega) = [W_1(\omega) \cdots W_M(\omega)]^T$ represents the stacked vector of the filter coefficients.

3 Multi-channel Wiener filtering

In the following, we will consider the problem of estimating as desired signal the speech component $X_{m_0}(\omega)$ of the m_0-th microphone arbitrarily selected to be the reference microphone. The multi-channel Wiener filter (MWF) produces a minimum-mean-square-error (MMSE) estimate by minimizing the MSE cost function [6]

$$\xi(\mathbf{W}_{m_0}(\omega)) = \mathscr{E}\{\left|X_{m_0}(\omega) - \mathbf{W}_{m_0}^H(\omega)\mathbf{Y}(\omega)\right|^2\}, \quad (4)$$

where $\mathscr{E}\{.\}$ denotes the expected value operator. The solution of this minimization problem is given by

$$\mathbf{W}_{m_0}(\omega) = \mathbf{\Phi}_y^{-1}(\omega)\mathbf{\Phi}_x(\omega)\mathbf{e}_{m_0}, \quad (5)$$

with $\mathbf{\Phi}_y(\omega) = \mathscr{E}\{\mathbf{Y}(\omega)\mathbf{Y}^H(\omega)\}$, $\mathbf{\Phi}_x(\omega) = \mathscr{E}\{\mathbf{X}(\omega)\mathbf{X}^H(\omega)\}$ the speech correlation matrix and $\mathbf{e}_{m_0}$ an M-dimensional vector with the m_0-th element equal to 1 and all other elements equal to 0, which selects the column of $\mathbf{\Phi}_x(\omega)$ corresponding to the reference microphone $m_0 \in \{1 \dots M\}$. The output signal can hence be written as

$$\begin{aligned} Z_{m_0}(\omega) &= \mathbf{W}_{m_0}^H(\omega)\mathbf{X}(\omega) + \mathbf{W}_{m_0}^H(\omega)\mathbf{V}(\omega) \quad (6)\\ &= Z_{x_{m_0}}(\omega) + Z_{v_{m_0}}(\omega), \end{aligned}$$

where $Z_{x_{m_0}}(\omega)$ corresponds to the speech component in the output signal and $Z_{v_{m_0}}(\omega)$ to the residual noise. Assuming that the speech and the noise components are uncorrelated, the correlation matrix $\mathbf{\Phi}_y(\omega)$ can be expressed as

$$\mathbf{\Phi}_y(\omega) = \mathbf{\Phi}_x(\omega) + \mathbf{\Phi}_v(\omega), \quad (7)$$

where $\mathbf{\Phi}_v(\omega)$ represents the noise correlation matrix, i.e., $\mathbf{\Phi}_v(\omega) = \mathscr{E}\{\mathbf{V}(\omega)\mathbf{V}^H(\omega)\}$.

To evaluate the performance of the MWF, we consider for each frequency bin the input SNR of the reference microphone, i.e.,

$$\mathrm{SNR}_{\mathrm{in}}^{m_0}(\omega) = \frac{\mathscr{E}\{\left|X_{m_0}(\omega)\right|^2\}}{\mathscr{E}\{\left|V_{m_0}(\omega)\right|^2\}} = \frac{\mathbf{e}_{m_0}^H\mathbf{\Phi}_x(\omega)\mathbf{e}_{m_0}}{\mathbf{e}_{m_0}^H\mathbf{\Phi}_v(\omega)\mathbf{e}_{m_0}}. \quad (8)$$

We also define the intelligibility weighted broadband input SNR which is obtained by weighting and integrating $\mathrm{SNR}_{\mathrm{in}}^{m_0}(\omega)$ over the full frequency band, i.e.,

$$\mathrm{SNR}_{\mathrm{in_{Br}}}^{m_0} = \sum_{\omega} I(\omega)\mathrm{SNR}_{\mathrm{in}}^{m_0}(\omega), \quad (9)$$

where the weight $I(\omega)$ expresses the importance of each frequency bin for speech intelligibility.

Similarly to the input SNR, we consider for each frequency bin the output SNR which can be computed as

$$\mathrm{SNR}_{\mathrm{out}}^{m_0}(\omega) = \frac{\mathscr{E}\{\left|Z_{x_{m_0}}(\omega)\right|^2\}}{\mathscr{E}\{\left|Z_{v_{m_0}}(\omega)\right|^2\}} = \frac{\mathbf{W}_{m_0}^H(\omega)\mathbf{\Phi}_x(\omega)\mathbf{W}_{m_0}(\omega)}{\mathbf{W}_{m_0}^H(\omega)\mathbf{\Phi}_v(\omega)\mathbf{W}_{m_0}(\omega)}, \quad (10)$$

and also define the intelligibility weighted broadband output SNR as

$$\mathrm{SNR}_{\mathrm{out_{Br}}}^{m_0} = \sum_{\omega} I(\omega)\mathrm{SNR}_{\mathrm{out}}^{m_0}(\omega). \quad (11)$$

4 Reference microphone selection

In this section we propose different procedures for reference microphone selection. First, the optimal reference microphone selection based on the broadband and the frequency-dependent output SNR are considered. We then present suboptimal procedures based on the input SNR, the signal energy and the distance.

4.1 Optimal reference selection

The optimal reference microphone selection scheme is obviously defined as the one resulting in the highest output SNR. Therefore, the reference microphone can be selected by first computing all possible Wiener filters $\mathbf{W}_m(\omega)$, $m = 1 \dots M$ and selecting as the reference microphone the microphone m_0 corresponding to the filter $\mathbf{W}_{m_0}(\omega)$ providing the highest broadband output SNR, i.e.,

$$\max_{m_0} \ \mathrm{SNR}_{\mathrm{out_{Br}}}^{m_0}. \quad (12)$$

As can be seen from (5), the MWF can be computed for each frequency bin separately and for each frequency bin, the output SNR $\mathrm{SNR}_{\mathrm{out}}^{m_0}(\omega)$ can be quite different for different reference microphones. Hence, we propose to further increase performance by optimizing the frequency-dependent output SNR, i.e. by selecting the reference microphone for each frequency bin individually.

The broadband and the frequency-dependent reference microphone selection based on the output SNR correspond to the optimal reference microphone selection procedures. However, the algorithm complexity increases since one first needs to compute M multi-channel Wiener filters. To avoid this drawback, we also consider different suboptimal selection procedures with a lower computational complexity.

4.2 Suboptimal reference selection

Since the output SNR of the MWF is generally closely related to the input SNR of the reference microphone, it is intuitive to estimate the desired signal component in the microphone with the highest input SNR, i.e.,

$$\max_{m_0} \ \mathrm{SNR}_{\mathrm{in_{Br}}}^{m_0}. \quad (13)$$

Furthermore, the computational complexity of this selection procedure by estimating the input SNR is much lower than by first computing M multi-channel Wiener filters prior to the optimal reference selection.

In theory, the input SNR dependents on the position of the desired source. The closer the source, the higher the input SNR and hence, if the distance of the desired source to all microphones is known, reference selection can also be performed by using the microphone closest to the desired source, i.e.,

$$\min_{m_0} d_{m_0}, \quad (14)$$

where d_{m_0} is the distance of the desired source to the m_0 microphone.

Another procedure we also propose is based on the broadband signal energy. Depending on the acoustical scenario [1], the selection of the microphone with the highest broadband signal energy can also be used as a suboptimal reference selection procedure, i.e.,

$$\max_{m_0} \ \sum_{\omega} \mathbf{e}_{m_0}^H\mathbf{\Phi}_y(\omega)\mathbf{e}_{m_0}. \quad (15)$$

[1] Would probably not work when the noise source is close to the microphone.

Similarly to the optimal procedure based on the output SNR, the reference microphone can also be selected for each frequency bin by using the suboptimal reference selection procedures based on the input SNR and the signal energy of the microphones. For each frequency bin the microphone with the highest input SNR or the highest signal energy is selected as the reference microphone, i.e.,

$$\max_{m_0} \frac{\mathbf{e}_{m_0}^H \mathbf{\Phi}_x(\omega)\mathbf{e}_{m_0}}{\mathbf{e}_{m_0}^H \mathbf{\Phi}_v(\omega)\mathbf{e}_{m_0}}, \text{ and } \max_{\mathrm{m_0}} \mathbf{e}_{\mathrm{m_0}}^{\mathrm{H}} \mathbf{\Phi}_{\mathrm{y}}(\omega)\mathbf{e}_{\mathrm{m_0}}. \quad (16)$$

5 Experimental results

In this section we investigate the performance of the MWF using the different reference selection procedures for a realistic acoustic scenario.

5.1 Setup and performance measures

Simulations have been performed using the acoustic scenario depicted in Figure 2. The circles (# 1…6) represent the microphone positions and the cross markers represent various positions of the desired source in a room with dimensions 7m×5m×3.5m and $T_{60} = 400$ms. We consider a scenario with a single speech source, and diffuse noise generated using the method introduced in [11]. The desired signal has been generated by convolving a clean speech signal from the HINT-database [9] with the impulse responses simulated using the image model [10]. The sampling frequency is $f_s = 16$kHz. For each position of the desired source, the closest microphone is assumed to be known and the input SNR is set to 5 dB for a source-microphone distance of 1.13 m.

In our implementation, we use the overlap/add method with a Hanning analysis and synthesis window, and we apply a 75% overlap between the signal frames. The FFT size used for the overlapp/add method is equal to $N_{\mathrm{FFT}} = 1024$. For the estimation of the correlation matrices, we use a perfect voice activity detector (VAD) to classify signal frames as speech dominant frames or noise dominant frames (silence). The correlation matrices $\mathbf{\Phi}_y(\omega)$ and $\mathbf{\Phi}_v(\omega)$ are estimated in a batch mode by using all speech + noise frames and all noise only frames respectively, i.e.,

$$\hat{\mathbf{\Phi}}_y(\omega) = \frac{1}{F_x}\sum_{F_x} \mathbf{Y}(\omega)\mathbf{Y}^H(\omega), \quad (17)$$

$$\hat{\mathbf{\Phi}}_v(\omega) = \frac{1}{F_v}\sum_{F_v} \mathbf{V}(\omega)\mathbf{V}^H(\omega), \quad (18)$$

where F_x and F_v are the number of frames during periods of speech + noise and periods of noise only. Since we assume that the speech and the noise components are uncorrelated, we estimate the speech correlation matrix $\hat{\mathbf{\Phi}}_x(\omega)$ as $\hat{\mathbf{\Phi}}_x(\omega) = \hat{\mathbf{\Phi}}_y(\omega) - \hat{\mathbf{\Phi}}_v(\omega)$. The resulting Wiener filter is computed as

$$\hat{\mathbf{W}}_{m_0}(\omega) = \hat{\mathbf{\Phi}}_y^{-1}(\omega)\hat{\mathbf{\Phi}}_x(\omega)\mathbf{e}_{m_0}. \quad (19)$$

In order to describe the MWF performance for all positions using a single number, we define the spatially averaged output SNR, i.e.,

$$\mathrm{SNR}_{\mathrm{out_{avg}}}^{m_0} = \frac{1}{\mathrm{N_s}}\sum_{\mathrm{N_s}} \mathrm{SNR}_{\mathrm{out_{Br}}}^{m_0},$$

which averages the intelligibility weighted broadband output SNR over the considered $\mathrm{N_s}$ source positions.

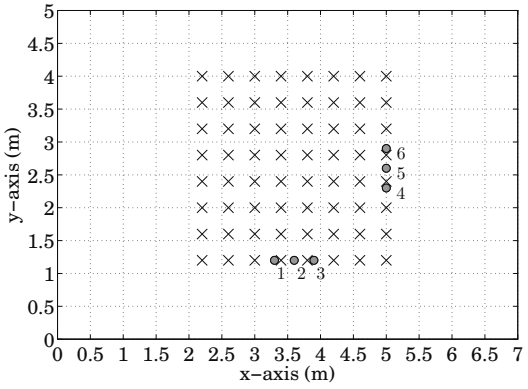

Figure 2: The scenario of an acoustic sensor network with $M = 6$ microphones.

5.2 Results

For a desired source located at the position with coordinates (4.2, 3.2), Table 1 shows the intelligibility weighted broadband output SNR of the MWF for all introduced reference microphone selection procedures. As one can see, the first microphone, arbitrarily selected as reference microphone, yields an output SNR which is 2dB smaller than the output SNR obtained using the sixth microphone as reference microphone, moreover showing the impact of the reference selection on performance in practical implementation of the MWF. From the results in Table 1, we observe that the presented reference selection procedures always lead to an improvement of the output SNR. However, the optimal reference microphone selection based on the output SNR increases the complexity of the algorithm and hence, suboptimal procedures have been investigated. The results clearly show that the less complex suboptimal procedures based on input SNR, energy and distance achieve similar performance as the optimal reference selection procedures based on the output SNR.

Figure 3 shows the intelligibility weighted broadband output SNR for different positions of the desired source by arbitrarily selecting the first microphone as the reference microphone. As expected, the choice of the first microphone as reference leads to good results at some positions of the desired source but to poor results at other positions. For example, a relatively small output SNR is achieved when the speaker is located in the area close to the microphones 4 to 6.

Figure 4 shows the output SNR when the reference microphone is selected by using the microphone which provides the highest broadband output SNR. Compared to the case when the first microphone is selected as reference, a higher or equal output SNR is obtained at all positions. Thus, using this broadband selection procedure, the optimal reference microphone is always selected and significant improvement in output SNR is obtained.

In Table 2 we compare the spatially averaged output SNR for all introduced reference microphone selection procedures. As expected, an arbitrary selected reference microphone (in this case the first microphone) yields poor performance. Similarly to the results in Table 1, we observe that all presented reference selection procedures also lead to an improvement of the spatially averaged output SNR. Moreover, the frequency-dependent procedures further increase performance compared to the broadband reference selection procedures.

$m_0 = 1$	$m_0 = 6$	**Broadband procedures**				**Frequency-dependent procedures**		
-	-	Output SNR	Input SNR	Energy	Distance	Output SNR	Input SNR	Energy
4.60 dB	6.5 dB	6.5 dB	6.5 dB	6.5 dB	6.5 dB	7.43 dB	7.11 dB	7.14 dB

Table 1: Output SNR for a desired source located at the position with coordinates (4.2, 3.2).

$m_0 = 1$	**Broadband procedures**				**Frequency-dependent procedures**		
-	Output SNR	Input SNR	Energy	Distance	Output SNR	Input SNR	Energy
6 dB	7.44 dB	7.41 dB	7.41 dB	7.37 dB	8.52 dB	8.23 dB	8.19 dB

Table 2: Output SNR, averaged over all considered source positions.

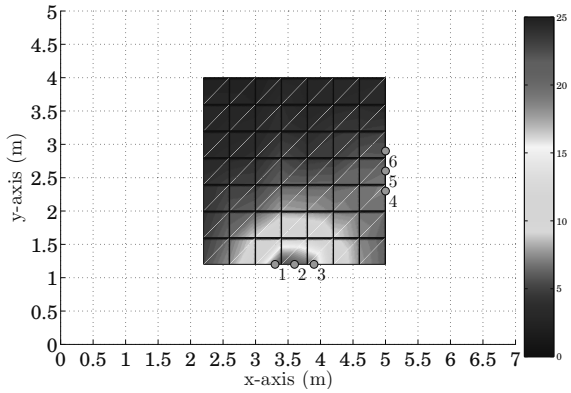

Figure 3: Position dependent output SNR obtained by using the first microphone as reference.

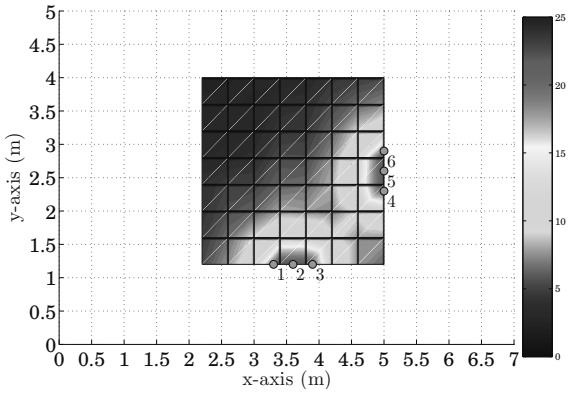

Figure 4: Position dependent output SNR with reference selection based on broadband output SNR.

6 Conclusion

In this paper, the performance of the MWF in an acoustic sensor network has been analyzed as a function of the selected reference microphone. Different optimal and suboptimal (broadband and frequency-dependent) reference selection procedures have been presented. It has been shown that compared to an arbitrary selected reference microphone, the broadband reference selection procedures lead to better performance even by using suboptimal procedures based on the input SNR and on the input energy. The broadband procedures have been extended to frequency-dependent procedures which further increase the output SNR.

References

[1] S. Srinivasan, "Using a remote wireless microphone for speech enhancement in non-stationary noise," *in Proc. ICASSP*, pp. 5088–5091, Prague, Czech Republic, May 2011.

[2] S. Markovich Golan, S. Gannot, and I. Cohen, "Performance analysis of a randomly spaced wireless microphone array," *in Proc. ICASSP*, pp. 121–124, Prague, Czech Republic, May 2011.

[3] T. C. Lawin-Ore and S. Doclo, "Analysis of Rate Constraints for MWF-Based Noise Reduction in Acoustic Sensor Networks," *in Proc. ICASSP*, pp. 269–272, Prague, Czech Republic, May 2011.

[4] A. Bertrand and M. Moonen, "Efficient sensor subset selection and link failure response for linear MMSE signal estimation in wireless sensor networks," *in Proc. EUSIPCO*, pp. 1092–1096, Aalborg, Denmark, Aug. 2010.

[5] 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.

[6] 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.

[7] 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.

[8] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multi-channel Wiener filtering and Generalized Sidelobe Cancellation for multi-microphone noise reduction in hearing aid applications," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 4, pp. 487–503, Jul. 2005.

[9] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

[10] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[11] E. A. P. Habets and S. Gannot,, "Generating sensor signals in isotropic noise fields," *Journal of the Acoustical Society of America*, vol. 122, no. 6, pp. 3464–3470, Dec. 2007.